

**Participatory Epidemiology: Harnessing the HealthMap Platform for
Community-Based Disease Outbreak Monitoring**

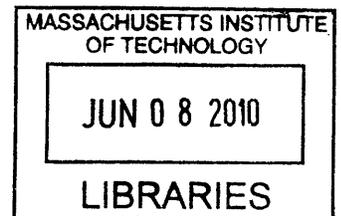
By Clark C. Freifeld

Bachelor of Science, Computer Science and Mathematics,
Yale University, 2000

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Media Arts and Sciences
at the
Massachusetts Institute of Technology

ARCHIVES



June 2010

© 2010 Massachusetts Institute of Technology
All rights reserved.

Signature of Author: _____
Program in Media Arts and Sciences
May 7, 2010

Certified by: _____
Franklin H. Moss, Ph.D.
Professor of the Practice of Media Arts and Sciences
Director, MIT Media Laboratory
Thesis Supervisor

Accepted by: _____
Professor Pattie Maes
Associate Academic Head
Program in Media Arts and Sciences

**Participatory Epidemiology: Harnessing the HealthMap Platform for
Community-Based Disease Outbreak Monitoring**

by Clark C. Freifeld

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning, on May 7, 2010
in Partial Fulfillment of the Requirements for the Degree of Master of Science in
Media Arts and Sciences



Read by: _____

Pattie Maes, Ph.D.
Associate Professor of Media Arts and Sciences, Program in Media Arts and Sciences
Thesis Reader

**Participatory Epidemiology: Harnessing the HealthMap Platform for
Community-Based Disease Outbreak Monitoring**

by Clark C. Freifeld

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning, on May 7, 2010
in Partial Fulfillment of the Requirements for the Degree of Master of Science in
Media Arts and Sciences

Read by: _____



John S. Brownstein, Ph.D.
Assistant Professor of Pediatrics, Harvard Medical School
Thesis Reader

**Participatory Epidemiology: Harnessing the HealthMap Platform for
Community-Based Disease Outbreak Monitoring**

by Clark C. Freifeld

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning, on May 7, 2010
in Partial Fulfillment of the Requirements for the Degree of Master of Science in
Media Arts and Sciences



Read by: _____

Chris Schmandt, M.S.
Principal Research Scientist, MIT Media Laboratory
Thesis Reader

Participatory Epidemiology: Harnessing the HealthMap Platform for Community-Based Disease Outbreak Monitoring

by Clark C. Freifeld

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning, on May 7, 2010
in Partial Fulfillment of the Requirements for the Degree of Master of Science in
Media Arts and Sciences

Abstract

Due to increasing global trade and travel along with a range of environmental factors, emerging infectious diseases such as Severe Acute Respiratory Syndrome (SARS), drug-resistant tuberculosis, and 2009 H1N1 continue to have significant impact on morbidity, mortality, and commerce worldwide. Early warning and detection of outbreaks plays an important role in protecting against disease, allowing public health authorities, organizations, and citizens to implement control measures rapidly. Due to its global reach, 2009 pandemic H1N1 represented not only a unique call to action for disease outbreak detection systems, but also precipitated increased public awareness of issues of emerging infection. This thesis explores the use of informal, user-contributed disease reports from the general public as a means to improve knowledge of local events and enhance early warning during the first and second waves of 2009 H1N1.

Building on the established HealthMap system, which has shown the effectiveness of using news media sources for rapid detection of outbreak events, we introduced the concept of “participatory epidemiology.” Through a series of software tools for Web and smartphone, we invited users from the general public to contribute their own knowledge and awareness of local activity. We deployed the system in two phases: in the first phase, users could contribute links to existing sources of online information; in the second phase, users could also contribute free-form reports of their own experiences or events in their local communities.

We received over 3,000 user submissions over the course of the study period from March 2009 to April 2010. We evaluated the system by examining a subset of notable reports and analyzing their timeliness as compared to previously existing HealthMap sources, as well as a range of qualitative factors demonstrating the potential for our approach. We further evaluated submissions relating to H1N1 in the U.S. by aggregating and comparing their volume to the Centers for Disease Control and Prevention influenza activity metrics, finding a Pearson’s correlation of 0.74.

Overall, the study indicates that with the appropriate tools, everyday citizens can play an important role in identifying and reporting infectious disease activity. The system is currently in active use and further development is ongoing.

Thesis Supervisor: Franklin H. Moss, Ph.D.

Title: Professor of the Practice of Media Arts and Sciences
and Director, Media Laboratory

Table of Contents

Abstract.....	3
1 Introduction.....	13
2 Background.....	16
2.1 Traditional Public Health Reporting Versus New Media Surveillance.....	16
2.2 The Emergence and Spread of 2009 H1N1 Influenza.....	21
2.3 Internet Tools for Collective Intelligence.....	25
2.4 Crowdsourcing for Health Applications.....	27
3 Problem and Approach.....	29
3.1 Limitations of the Established HealthMap System.....	29
3.2 Approach.....	30
3.3 Participatory Epidemiology Can Work Today.....	31
4 Software Tools and Data Collection Methods.....	33
4.1 Software Tools.....	33
4.2 URL-based Report Acquisition.....	33
4.3 Smartphone Applications.....	35
4.4 Voice and Web Applications.....	40
4.5 Clinical Reporting Tool.....	41
4.6 Submission Review and Validation Methods.....	43
5 Community Response.....	46
5.1 Software Tools for the General Public.....	46
5.2 Response in the Professional Public Health Community.....	48
6 Results.....	50
6.1 Overview of Data Collected.....	50
6.2 Spam and Otherwise Unhelpful Reports.....	57
6.3 Analysis of Selected Sample Submissions.....	57
6.3.1 News-Based Submission via Mobile.....	57
6.3.2 Local Clusters and Personal Accounts.....	59
6.3.3 Closures and School-Based Outbreaks.....	60
6.3.4 Clinical Reports.....	61
7 Discussion and Limitations.....	62
8 Conclusions and Future Work.....	64
9 References.....	66
Acknowledgements.....	70
Appendix A. Review of Selected H1N1 Applications for iPhone and Android.....	71
Appendix B. Outbreaks Near Me Is Top Free App in Healthcare & Fitness.....	81
Appendix C. End-user License Agreement for Submissions to HealthMap.....	81
Appendix D. Selected User-Submitted Photographs.....	83

1. Introduction

Infectious disease outbreaks continue to have significant impact on morbidity, mortality and commerce worldwide. The emergence of diseases such as West Nile Virus, Severe Acute Respiratory Syndrome (SARS), and drug-resistant tuberculosis are just a few examples of recent outbreak events with global effects, despite advances in antibacterial and antiviral treatments. However, the appearance of novel H1N1 influenza (“swine flu”) in spring 2009 represents the first emerging infectious pandemic of global scale in recent memory. Due to its impact across borders and across social strata, H1N1 represents a unique call to action for modern disease outbreak detection and response systems, as well as a transformation of worldwide cultural consciousness of infectious disease threats.

Because disease can spread exponentially within a population, and due to the reach and volume of international travel today, a new infection can rapidly attain global scale. Early detection of outbreaks is therefore an essential weapon in the fight against emerging pathogens, allowing us to implement control measures rapidly, to treat those infected, and to prevent further spread.

Part of a new generation of online detection tools, HealthMap is a multi-lingual, real-time disease outbreak tracking and visualization Web site [1]; launched in fall 2006, the system collects 600 reports per day in seven languages, from both general news media and public health sources around the world. Updated hourly, the system filters these reports to determine relevance, disease, location, and duplication clustering by means of a series of automated text processing algorithms. Relevant reports are then aggregated and displayed in a freely available dashboard where users can tailor the view according to date, disease, location and source. HealthMap provides an overview of real-

time information on emerging infectious diseases and has particular interest for public health officials, health practitioners, international travelers and the general public.

Building on the HealthMap system, the work of this master's thesis project consisted of improving disease outbreak surveillance by fostering user-driven reporting. Under the established HealthMap system, news media sources, when filtered effectively, have proven their worth for disease outbreak detection and early warning, despite their large volume and lack of official verification. However, with this project, we focus on yet more informal, unofficial, and unverified sources of disease outbreak information: everyday users. As demonstrated by the success of sites such as Wikipedia [2], "crowdsourcing" projects have proven that harnessing everyday volunteers can produce effective outputs, even surpassing the efforts of paid experts in some cases. For this particular application, the hypothesis is that while allowing any user to make a report may represent a decrease in reliability as compared with other sources, it represents an increase in timeliness and sensitivity. While government public health agencies and news media organizations may have better tools and resources for verifying local events, they lack the highly distributed, ubiquitous network of "eyes on the ground" that is possible through citizen participation.

Moreover, in addition to simply improving the output of the HealthMap system, by engaging the public in the practice of epidemiology, we aim to transform users from passive recipients of information to active participants in a collaborative community, to reimagine the role of ordinary citizens in global public health. The analysis of the usefulness and value of the user-contributed data that we carry out with this thesis project is not only a measure of the validity of user sources as a concept, but also a test of our own ability to provide the tools necessary to engage users, tools that are seamless, rewarding to use, and foster a vibrant community. Further, at the current moment, we

are presented with a unique historic opportunity to deploy and test this project. Thanks to the established success and high visibility of the existing HealthMap system, combined with new popular awareness of emerging disease threats due to H1N1, ordinary users are uniquely prepared for engagement in citizen reporting.

In the sections that follow, we will present the Background for the project, outline the Problem and Approach, describe Tools and Methods in detail, summarize Community Response to the project, report Results of the study, and finally offer Discussion and Conclusions as to the broader implications of the work and possible future directions.

2. Background

2.1 Traditional Public Health Reporting Versus New Media

Surveillance

Traditional public health reporting systems generally consist of a hierarchical structure of federal, state, and local authorities interacting with clinical providers. In general, it is incumbent upon the clinician to actively report infection information to local health authorities. In many cases, reporting occurs by mandate, as the government maintains a list of specific diagnoses that must be reported [3]. However, often a definitive diagnosis is not possible at the initial presentation of the patient, thus introducing a delay in the reporting process. Especially in the case of an unknown novel disease, it is not possible to establish reporting mandates in advance, and the system is dependent upon the judgment of the “astute clinician” to recognize and report cases that seem unusual or otherwise noteworthy to public health authorities [4]. While many infectious disease specialists are constantly on alert for any suspicious symptoms in their patients, many clinicians, both in the U.S. and around the world, are continually overburdened with meeting basic obligations for patient care and understandably have limited time for reporting.

By contrast, broad adoption of the Internet as a tool for rapid and low-cost dissemination of information has enabled new, less formal modes of public health reporting. Rather than a top-down hierarchy, the Internet allows for instantaneous broadcast communications and unencumbered information flow between agencies and individuals in all directions. New communication tools such as email and the World Wide Web thus present important opportunities to address the limitations of traditional public health reporting. However, the massive volumes of information now generated create a significant risk of information overload, leaving public health officials and

responders overwhelmed with reports and alerts, unable to find relevant information and unable to know how and when to act.

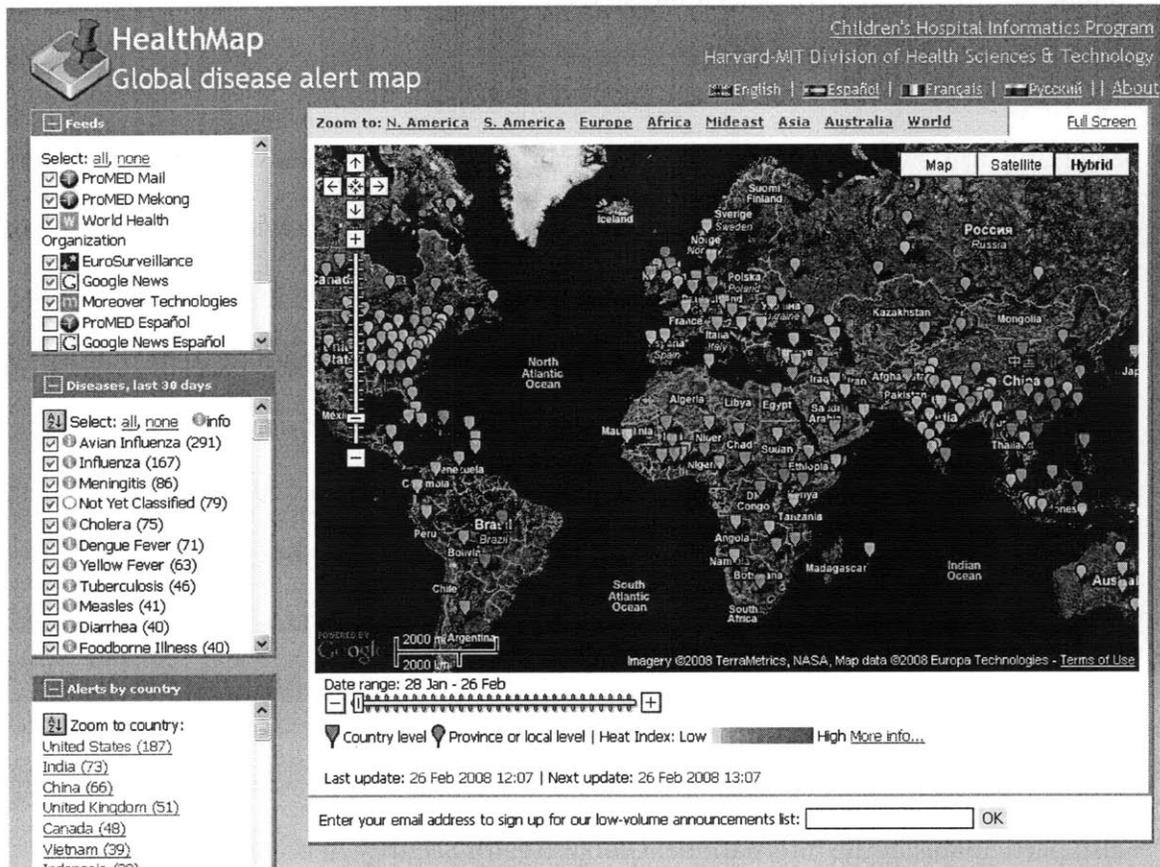


Figure 1. The HealthMap real-time global disease dashboard. The user-friendly interface allows the site visitor to rapidly filter and navigate a large volume of outbreak information.

As an approach to the challenge of how to harness the massive array of informal information sources to create a useful situational awareness tool for public health, we created the HealthMap system in 2006 (Figure 1). As noted above, the system uses custom-designed, automated artificial intelligence algorithms to collect, filter, and classify outbreak reports on a 24-hour basis. HealthMap in turn provides a sophisticated Web-based visualization dashboard to allow both public health officials and ordinary laypeople rapid and easy access to real-time disease information. Less than a month after its launch, it was featured in a news article on Wired.com and rapidly gained users,

recognition, and support in the ensuing years. At the time of this writing in April 2010, the site receives over 150,000 visits per month from international, nation, and local public health officials as well as interested citizens around the world.

With the HealthMap system, we have found that in many cases, indications of emerging disease outbreaks often appear in informal sources in advance of official reporting channels. For example, we performed a retrospective review of avian influenza in both animals and humans, using detailed case documentation by the World Health Organization, and compared the date of the first official report with the date of the first report collected by the HealthMap system. For the full calendar year 2007, comprising roughly 100 outbreaks, on average the HealthMap reports preceded the official reports by 2.3 days (Figure 2). In 2008, when food-borne Salmonella infection struck forty U.S. states, through informal HealthMap reports, we were similarly able to track both first signs and the progress of the outbreak, ahead of official reports (Figure 3).

As one of the leading global online disease surveillance systems [5], HealthMap also participates in the international Global Health Security Action Group (GHSAG) consortium, a collaborative initiative of the G7 countries, focused on early detection of chemical, biological, nuclear and radioactive events worldwide.

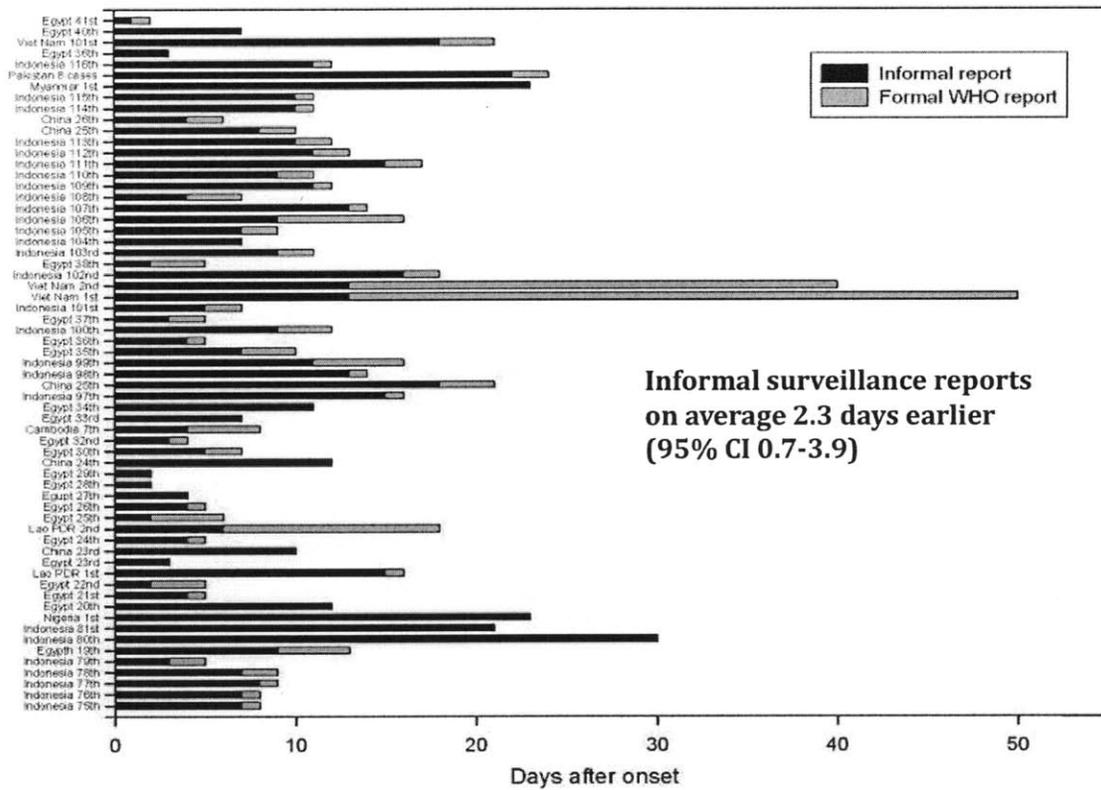


Figure 2. Time differentials between HealthMap informal reports and official WHO announcements of avian influenza events in 2007. Date of onset of the disease in each event was determined through the official WHO case history.

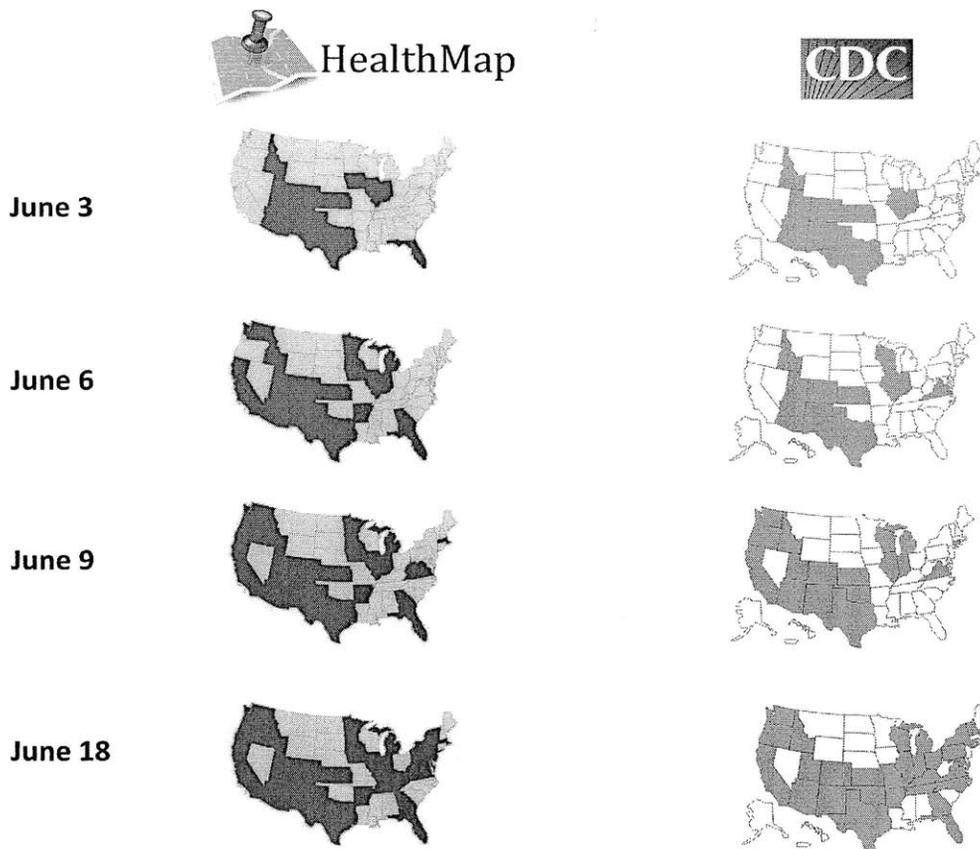


Figure 3. Maps of the spread of Salmonella Saintpaul in the U.S. in 2008. The data collected through informal sources by the HealthMap system closely mirror and in some cases anticipate the officially confirmed reports from the CDC.

Effective pan-national public health awareness tools are becoming only more important in the modern era, as novel emerging disease events have continued to increase decade by decade [6], as shown in Figure 4, even before the appearance of 2009 H1N1. The reasons that infectious disease threats are on the rise despite rapid advances in antibiotics, antiviral medications, infection control measures and other treatments are varied, complex and by no means fully understood. However, some key factors include increasing population density, especially in cities all over the world; greater animal-human contact through increasing human encroachment on previously undisturbed animal habitats; greater animal-human contact through live animal markets in densely

populated cities, particularly in Asia; greater volume of high-speed international travel; and climate change. All of these factors make the current time an especially apt moment to explore new approaches to infectious disease detection and monitoring.

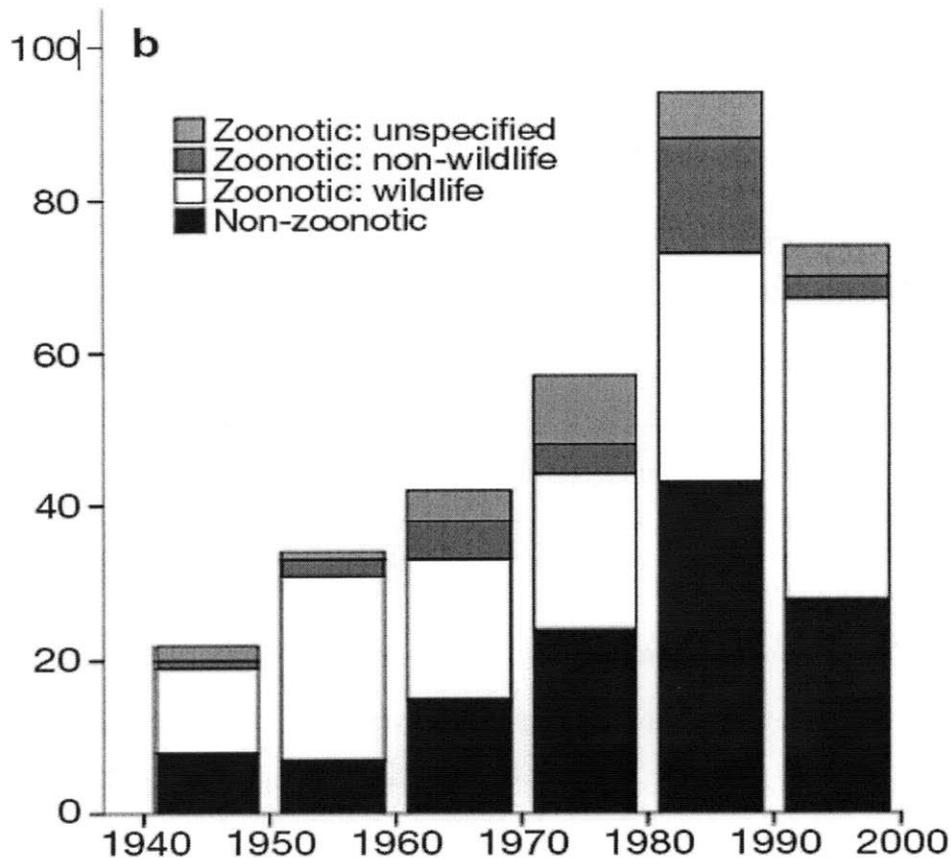


Figure 4. Counts of emerging infectious disease events by decade, from [6], showing a strong upward trend as well as the large proportion of zoonotic events.

2.2 The Emergence and Spread of 2009 H1N1 Influenza

A majority of recent novel disease outbreaks are of zoonotic pathogens [6], namely, infections that cross the species barrier from animals to humans. 2009 H1N1 was no exception to this trend: it is believed to have first passed from pigs to humans at a pig farm in the village of La Gloria, Veracruz, Mexico in March 2009 [7], possibly as

early as February 2009 [8], and achieved human-to-human transmission shortly thereafter. The first signal of the outbreak to appear in the HealthMap system came on April 1, in the form of a Spanish-language news article from local newspaper La Jornada, reporting that as many as 60% of La Gloria's 3,000 inhabitants had suffered from an unidentified respiratory illness [9]. Two pediatric deaths from respiratory illness were also reported in the same article, although it was not known whether the deaths were linked to the more widespread outbreak. Although the information contained in the news report itself was not enough to warrant an international alarm, it does serve as an important reference point for our retrospective understanding of the course of the pandemic, and a demonstration of how such signals appear in the HealthMap system.

The first two known U.S. cases of 2009 H1N1, with onset of symptoms in late March, occurred near San Diego, California and were first identified by the Centers for Disease Control and Prevention (CDC) on April 13 and 17, and published by the CDC on April 21 [10]. The CDC analysis was among the first genetic laboratory analyses to establish concretely that these unusual respiratory symptoms indeed resulted from a novel swine-origin influenza virus. The first announcement of the World Health Organization (WHO) appeared just days later, on Friday, April 24 [11]. Meanwhile, on the same day, traffic to the HealthMap Web site doubled from its baseline of approximately 2,000 visits per day to 4,000 visits. The following day, a Saturday, I returned from an afternoon bike ride with my family to discover that the HealthMap site was down, the database overloaded by too many connections. I quickly restored the server and thus began several days of often feverish, around-the-clock work to patch the software, improve caching, manage disk space, and boost the performance of the system in order to stay online. On Sunday the 26th, HealthMap was featured on Cable Network News (CNN) as part of its ongoing reporting on the emergence of H1N1. Over the

following four days, the flow of disease outbreak reports into the system increased ten-fold even as the visit traffic surged by two orders of magnitude to over 200,000 visits in a single day.

Meanwhile, due to international travel, cases and deaths rapidly began to appear around the world. From initial cases in Mexico, the U.S. and Canada, the virus spread to the UK, France, and Israel [12]. For the first time in history, the WHO raised its pandemic alert level to 5, indicating human-to-human spread in at least two countries and a signal of imminent pandemic [13]. With the HealthMap system, we first created a special visualization page dedicated to H1N1 reports along with reports of unidentified respiratory illness suspected to be linked to the new influenza. In parallel, we began to identify and collect case-level data for the first time, annotating reports with counts of ruled-out cases, suspected and confirmed cases, as well as suspected and confirmed deaths, from both official reports (including the CDC, WHO and Public Health Agency of Canada) and unofficial news media reports. On May 5th, we launched the HealthMap H1N1 visualization page (<http://healthmap.org/h1n1>) showing case counts and an animated map view depicting the spread of the disease, as part of the New England Journal of Medicine's H1N1 Influenza Center [14].

One month prior to the initial recognition of H1N1, in mid-March 2009, as part of the current thesis research, we deployed the capability for HealthMap visitors to submit news articles into the system, to contribute their own awareness to fill gaps in our coverage. As detailed below in Section 6.1, this capability proved critical for HealthMap during the initial recognition of H1N1.

After a brief summer lull, the virus returned in force with the arrival of the fall season in North America. In the U.S., its impact began with school closures in late August and was in full force by mid-September. We released the HealthMap iPhone

application, Outbreaks Near Me, on September 1, allowing users to receive geo-targeted real-time alerts as well as to submit their own event reports into the system, thus officially launching our “participatory epidemiology” concept. The release of the application met with broad attention and acclaim from a range of media outlets including Wired Science [15], the Wall Street Journal [16], and the NBC Nightly News with Brian Williams [17]. Shortly after releasing the application, we also added the submission feature to the Web site and began collecting a steady flow of reports through both channels. Meanwhile, by September 26, the second wave had claimed 1,379 lives in the U.S. [18]. In order to help keep the public informed of ongoing influenza activity, the Department of Health and Human Services turned to HealthMap to provide a rich interactive visualization system for the flu.gov Web site (<http://www.flu.gov/wheretheyoulive/healthmap/>). This tool provides overlays of weekly CDC influenza tracking indicators, incorporates Twitter and YouTube updates from state health departments as well as real-time HealthMap alerts.

Figure 5 presents a summarized timeline of key H1N1 events both in the general progress of the pandemic and for the HealthMap system specifically. Suffice to say, the emergence of 2009 H1N1 served as the first major test of the HealthMap system and its capacity to provide early warning for public health officials as well as inform the general public on the progression of the pandemic. With a broad range of information choices available at their fingertips, including commercial media outlets as well as official reporting channels, government decision makers and concerned citizens together continually chose HealthMap as their preferred system for up-to-date information. The combination of HealthMap’s position as a global leader in real-time disease surveillance and the worldwide focus on influenza made 2009 and early 2010 a uniquely suited historical moment to carry out the current master’s thesis work, to study these phenomena and to provide a significant contribution to the field.

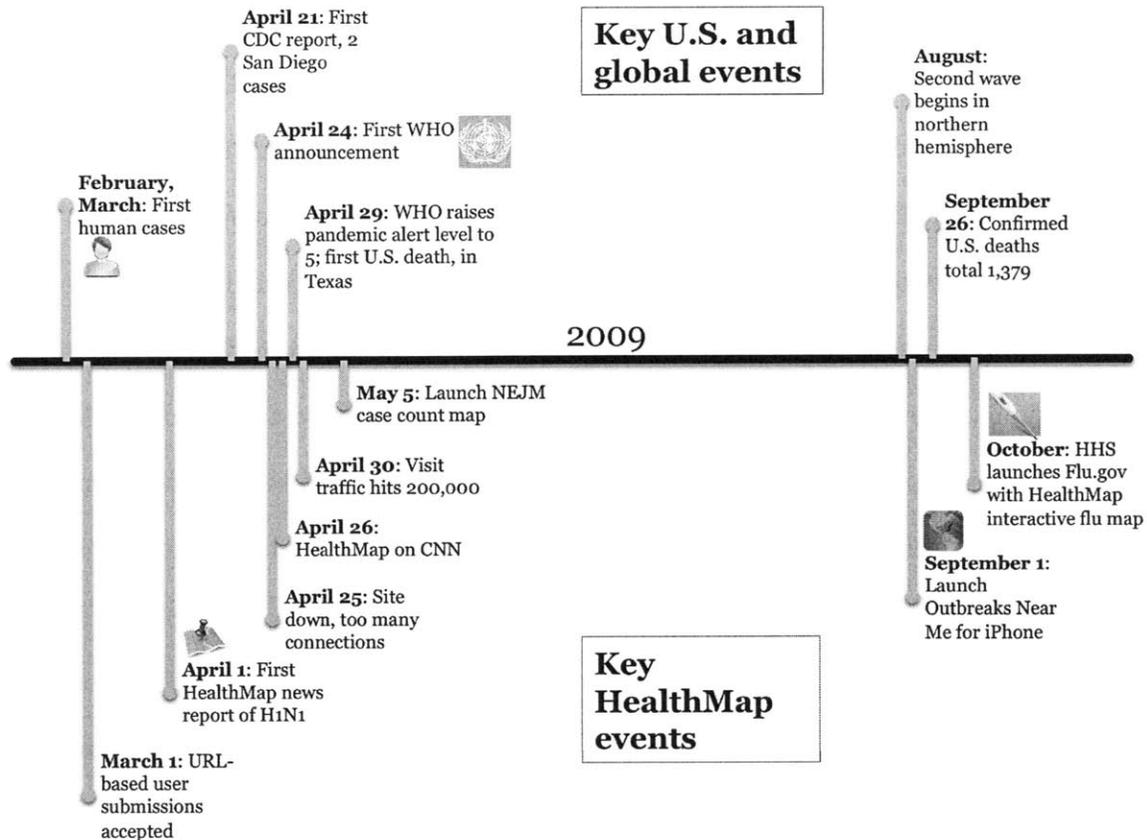


Figure 5. Key events in the progression of the 2009 H1N1 pandemic, both for global public health and for the HealthMap system.

2.3 Internet Tools for Collective Intelligence

Broad adoption of the Internet around the world has enabled a new class of so-called “collective intelligence” tools, systems that allow many people to share information and work together to solve problems and create products and services that were previously impossible [19]. Wikipedia [2], InnoCentive [20], Threadless [21], and the ESP Game [22] are just a few examples of how organizations and individuals have effectively harnessed the collective efforts of loosely associated volunteers to achieve unparalleled outcomes.

Collective intelligence applications fall into two key categories, passive and active. In the case of passive collective intelligence, a system makes use of the aggregated “side

effects” of individual activity to reveal new insights. For example, Amazon.com mines customer purchase histories in order to create new product recommendations [23,24]; Google maps the links between Web pages [25] in order to identify the most well-regarded sites; the ESP Game uses many thousands of game-play sessions to add semantic information to images. In these cases, the users who are contributing to the collective intelligence product may be unaware or only vaguely aware of their contribution as they pursue another specific short-term goal. The established HealthMap system is an example of this model: by integrating articles across a broad range local news sources, we create a comprehensive picture of ongoing global disease activity as well as detecting early signals of outbreaks before they spread. We accomplish this view despite the fact that each individual newspaper is working primarily to serve the needs of its local community by providing timely, informative reporting on local events, unaware of its role in any outbreak tracking system.

By contrast, in an active collective intelligence application, the participants are consciously working together toward a common goal, with Wikipedia [2] being perhaps the most large-scale and well-known currently active example. In this type of application, while users may have a diversity of purposes and reasons for participating in the community, at some level, they share the common vision and generally commit their time and resources altruistically, for the benefit of the community, without necessarily an immediate short-term individual benefit. With the work of this thesis, we are augmenting our existing epidemiology with *participatory* epidemiology, engaging users directly in contributing to the system. Rather than relying on professional journalists and commercial media to collect and disseminate event information, we allow any interested citizen to become a reporter.

2.4 Crowdsourcing for Health Applications

While the HealthMap participatory epidemiology initiative is the first effort of its kind and scale, a number of important projects have already begun collaborative efforts in the health field. Google Flu Trends [26,27] is a notable example of passive collective intelligence—it taps the variations in how Internet users search the Web as an indicator of flu activity for a given geographical area. Likewise, systems such as the Distribute [28] project of the International Society for Disease Surveillance and the Automated Epidemiologic Geotemporal Integrated Surveillance System (AEGIS) [29] both currently integrate emergency department visit records from a range of hospitals to provide real-time disease and syndromic surveillance.

In terms of more active engagement with users, the Program for Monitoring Emerging Diseases (ProMED) has run an email distribution list of infectious disease experts for over 15 years [30]. The list has tens of thousands of subscribers and employs a mixed model of top-down moderation and grassroots information discovery. In addition to the core board of moderators, all of whom are infectious disease experts, a distributed network of “rapporteurs” are recruited and often send news articles to the moderators for review, approval and subsequent posting to the distribution list. In a few cases, the “astute clinician” as described above in the traditional model of public health reporting has also been astute enough to be aware of ProMED and also contributed firsthand knowledge of a nascent outbreak, as occurred in the initial recognition of SARS in 2003 [31]. For the HealthMap project, ProMED is a key source of information and we have worked closely with the ProMED community, often providing moderators and rapporteurs with feeds of information from our automated system to supplement their ongoing work.

Finally, in the area of crisis mapping and response, the Ushahidi project has seen success with engaging users on the ground to provide first-person accounts of rapidly-developing crisis situations [32]. While public health is not the primary focus of Ushahidi, in many cases, portions of the information people contribute are highly relevant to public health, and in some cases directly relate to infectious disease outbreaks. Most notably, the Ushahidi platform and its users proved instrumental in mounting a response to the earthquake in Port-au-Prince, Haiti on January 12, 2010. The group is credited with saving hundreds of lives by aiding in locating those trapped under the rubble in the immediate aftermath of the quake [33]. The participants also generally helped responders such as the International Committee of the Red Cross and the U.S. military stay informed of events and conditions on the ground. In this particular effort, HealthMap collaborated with Ushahidi to help disseminate information coming from the Ushahidi platform, and provided a focused map for tracking infectious disease activity in Port-au-Prince and around Haiti (<http://healthmap.org/haiti>).

Overall, new paradigms in collective intelligence made possible through the Internet represent a promising area for application to a range of challenges in public health; the work of this thesis project applies these concepts to addresses the challenge of global disease monitoring in particular.

3. Problem and Approach

3.1 Limitations of the Established HealthMap System

Although we have had success with the original HealthMap approach of using automated natural language processing algorithms to filter and integrate a vast volume of electronic media, fundamental limitations remain. Due to inevitable limits on computing resources, skilled person-hours of expertise across languages, and the challenges of automated parsing of radio and video content, the signal-to-noise ratio is continually difficult to boost significantly simply by expanding on our original methodology.

As part of our ongoing work, and especially during the initial wave of H1N1 infections, we have implemented a range of performance enhancements to increase the scalability of the system. However, as it stands now, with the five algorithms we apply across seven languages (document body text extraction, disease and location extraction, Bayesian filtering for relevance, and duplication clustering), the processing overhead is significant given the volume of reports. Further, in order to train the machine-learning algorithms, we face significant upfront demands for painstaking work by human curators who must combine foreign language and infectious disease epidemiology skills with remarkable patience for tedious review and classification work.

The process becomes still more labor intensive if we consider the large volume of important non-text media, namely television and radio broadcasts, which would first need to be converted to text before entering the existing pipeline. Further, even as adoption of mobile phones is skyrocketing [34], much news media in developing countries is disseminated via analog radio broadcast [35], and thus largely inaccessible to any Internet search engine. At the same time that the current HealthMap approach is

limited by the lack of online news media sources in developing countries, these countries are often places with a relatively high burden of infectious disease, and known hotspots for novel emerging pathogens [6]. However, with mobile-capable direct citizen reporting, we can begin to address some of these limitations.

3.2 Approach

One of our key goals in developing the system was to enable users to easily and conveniently contribute reports into the system. In support of this goal, we have created a broad range of input channels for the user to choose from, as detailed below. Further, in general, it has been our philosophy to solicit relatively unstructured information from users, for two reasons. One is again to make contributing quick and easy, rather than burdening users with lengthy forms, such as those one might see in a typical outbreak investigation (Figure 6). Secondly, we are counting on the user for early detection of weak signals of previously unrecognized events. While we have some concept of how an outbreak event may appear, we want to enable the user to represent what he feels is important as regards disease activity in his community, even if it may not fit a predetermined model. Fortunately, with the existing HealthMap system, we have established expertise and algorithms for processing large volumes of unstructured data, so we can shift the burden away from the user and into the system.

Along the same lines, another key part of our approach philosophy is to empower users to participate in the public health process. One motivation for this thesis project is certainly to improve the ability of HealthMap to achieve its original purposes of early outbreak detection, and broad global coverage. However, rather than simply improving the system, with the current work, we also shift the focus. Our broader goal is now to engage users as active participants, rather than passive consumers of the information we collect. While many people may view protecting the population's health as the purview of

distant government agencies or academic research groups, by valuing local knowledge, we aim to shift this culture by creating a new and valuable role for the everyday person.

While with this approach, we certainly sacrifice the skills of trained journalists and the accountability structures already in place for news media, we gain access to citizens’ real-time knowledge and awareness of their local environment. Although the information will still come with noise, we also eliminate the noise inherent in “passive” collective intelligence because our sources will for the most part be conscious of the mission and share our goals.

Part I General information					
1. Type of disaster: <input type="checkbox"/> Hurricane (name _____) <input type="checkbox"/> Heat wave <input type="checkbox"/> Tornado <input type="checkbox"/> Technological disaster <input type="checkbox"/> Flood <input type="checkbox"/> Terrorism <input type="checkbox"/> Earthquake <input type="checkbox"/> Other (specify) _____			2. Facility type (info source): Please check one that best applies. <input type="checkbox"/> ME office <input type="checkbox"/> Funeral home <input type="checkbox"/> Nursing home <input type="checkbox"/> Coroner office <input type="checkbox"/> Hospital <input type="checkbox"/> DMORT office <input type="checkbox"/> Other (specify) _____		
3. Facility address: Street _____ County/parish _____ State _____ Z-code _____			4. Contact person (informant): Name _____ Phone number _____ Email Address _____		
Part II Deceased information					
5. Case / medical record number: _____			6. Body identified? <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Pending		
7. Date of Birth (MM/DD/YY) / / <input type="checkbox"/> Unknown			8. Age in years: <input type="checkbox"/> < 1 yr <input type="checkbox"/> Unknown		
9. Residential address of decedent: County/parish _____ City _____ State _____ Zip code _____		10. Ethnicity: <input type="checkbox"/> Hispanic <input type="checkbox"/> Non Hispanic <input type="checkbox"/> Unknown		11. Race: <input type="checkbox"/> American Indian or Alaskan Native <input type="checkbox"/> White <input type="checkbox"/> Black or African American <input type="checkbox"/> Asian <input type="checkbox"/> Native Hawaiian or other Pacific Islander <input type="checkbox"/> Other race	
12. Gender: <input type="checkbox"/> Male <input type="checkbox"/> Female <input type="checkbox"/> Undetermined	13. Date of Death: (MM/DD/YY) ____ / ____ / ____ <input type="checkbox"/> Unknown	14. Time of Death: ____ (24 hr clock) <input type="checkbox"/> Unknown	15. Date of body recovery: (MM/DD/YY) / / <input type="checkbox"/> Unknown		
16. Time of body recovery: <input type="checkbox"/> ____ (24 hr clock) <input type="checkbox"/> Unknown		17. Place of death or body recovery: <input type="checkbox"/> Decedent's home <input type="checkbox"/> Evacuation Center/shelter <input type="checkbox"/> Vehicle <input type="checkbox"/> Hospital <input type="checkbox"/> Hotel /motel <input type="checkbox"/> Nursing Home / long term care facility <input type="checkbox"/> Hospice facility <input type="checkbox"/> Unknown <input type="checkbox"/> Street/Road <input type="checkbox"/> Prison or detention center <input type="checkbox"/> Other (specify) _____			
18. Location of death or body recovery: State _____ county/parish _____ Intersection _____		19. Prior to death, the individual was a: <input type="checkbox"/> Resident <input type="checkbox"/> Non-resident-intrastate <input type="checkbox"/> Unknown <input type="checkbox"/> Foreign <input type="checkbox"/> Non-resident-interstate <input type="checkbox"/> Other			
20. Was the individual paid or volunteer worker involved in disaster response? <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown		21. Body recovered by: <input type="checkbox"/> Law enforcement <input type="checkbox"/> Fire department <input type="checkbox"/> DMORT <input type="checkbox"/> Other (specify) _____ <input type="checkbox"/> EMS <input type="checkbox"/> Search and rescue <input type="checkbox"/> Family or individual <input type="checkbox"/> Unknown			
Part III Cause and Circumstance of death (check one that best applies)					
22. Mechanism or cause of death— Injury <input type="checkbox"/> Drowning <input type="checkbox"/> Electrocution <input type="checkbox"/> Lightning <input type="checkbox"/> Motor Vehicle occupant/driver <input type="checkbox"/> Pedestrian/bicyclist struck by vehicle <input type="checkbox"/> Structural collapse <input type="checkbox"/> Fall <input type="checkbox"/> Cut/struck by object/tool		23. Cause of death— Illness <input type="checkbox"/> Neurological disorders <input type="checkbox"/> Meningitis/encephalitis <input type="checkbox"/> Seizure disorder <input type="checkbox"/> Stroke (hemorrhagic or thrombotic) <input type="checkbox"/> Other (specify) _____ <input type="checkbox"/> Respiratory failure <input type="checkbox"/> COPD <input type="checkbox"/> Pneumonia		24. Cause of death: <input type="checkbox"/> Confirmed <input type="checkbox"/> Probable <input type="checkbox"/> Pending <input type="checkbox"/> Unknown	
		25. Relationship of cause of death to disaster: <input type="checkbox"/> Direct <input type="checkbox"/> Possible <input type="checkbox"/> Indirect <input type="checkbox"/> Undetermined			
		26. Circumstance of death: (free text)			

Figure 6. A typical form for public health data collection. This particular form from the CDC is for post-disaster mortality tracking, <http://www.bt.cdc.gov/disasters/surveillance>.

3.3 Participatory Epidemiology Can Work Today

Finally, a key aspect of our approach is that we have all the elements in place to make the system work now. Other than perhaps broader adoption of smartphones, we

don't assume technologies that have yet to be invented or government mandates, funding, and laws that have yet to be adopted. We present not only a vision for empowering citizens to play a role in public health, but also the realization of this vision with actual citizens participating during a global pandemic. Although the work is a preliminary exploration, our ability to deploy it in the real world, across a large number of users, and maintain it on an ongoing basis, gives unique insight into how people respond to our approach and the true potential for the project. Finally, even after the work of this thesis is complete, all the tools and techniques described here will continue to run, serve users, collect data and be refined and improved.

4. Software Tools and Data Collection Methods

4.1 Software Tools

To accomplish our goal of engaging users and collecting real-time local information, we created and deployed a broad range of technical products. These included an ensemble of Web and mobile user interfaces as well as a variety of backend services to support the flow of information through the system.

4.2 URL-based Report Acquisition

Our initial foray into user-contributed content consisted of placing the “Outbreak missing from the map?” prompt on the HealthMap Web site and allowing anyone to submit the Uniform Resource Locator (URL) of a news article or other online information source, and have it posted to HealthMap. With the URL, the system loads the content from the original source, converts it to the UTF8 character set, parses the headline, automatically detects the language, and prepares the text for input to the named-entity extraction engine by removing any accents from accented characters. Originally, the system would also check the database of existing reports for duplicates and reject the submission if it determined that we had already collected an equivalent article; however, due to the challenge of matching accurately and the low volume of duplicate submissions, we ultimately removed the check in order to encourage participation and avoid the risk of incorrectly rejecting a legitimate contribution.

Once the report is loaded and parsed, if the user supplied an email address with her submission, the system then creates a temporary authentication token and generates an email to the contributor. The email contains a link with the token embedded in it, allowing the user access to the HealthMap administration tool, so that she can make edits to the report (Figure 7). One of the key concepts we strive to implement with this

4.3 Smartphone Applications

Our efforts to enlist users for public health surveillance began in earnest with the development and launch of the HealthMap iPhone application, Outbreaks Near Me. In addition to providing much of the same successful functionality of the HealthMap Web site in a smartphone form factor, the concept behind the application was two-fold. The first principle was to take advantage of the phone's built-in Global Positioning System (GPS) capability to provide the user with real-time HealthMap alerts relevant to his current location, including the ability to receive targeted "push" alerts immediately to the phone without needing to actively query the system. The second key idea was to allow users to submit free-form, firsthand accounts of any disease activity in their local community. With the release of the iPhone app, in addition to being able to submit third-party content such as news reports, users could report on events such as school closures and absenteeism, as well as illness in those around them or their own symptoms.

While the main focus of this thesis is the second idea, these two concepts were essential together: in order to motivate people to participate in the system, it was important to provide targeted alerts, so as to give immediate value to any user. Then, users who had initially installed the application just to receive information, could in turn give back to the community by sharing their own knowledge with others. Thus with this approach, we effectively leveraged the existing HealthMap tools to give momentum to the feedback cycle.

As depicted in Figure 8, the client-side user interface consists of five different screens: Map, List, Submit, Info and Settings. The Map view determines the phone's location coordinates and automatically submits them to the HealthMap backend Web service. From this screen the user can also perform explicit searches as desired, based on location, disease or keyword. The Web service (described in more detail below) then

returns a set of alerts for display on the map. The same alerts are also displayed on the List screen. The Settings screen allows the user to set parameters for receiving “push” alerts, such as which diseases and locations to target, and whether to send the alerts via email or directly to the phone via Apple’s push service. The Submit screen presents a simple form prompting the user to enter disease, location, and other basic information about the event she is reporting. The form also allows the user to take a photograph with the phone’s built-in camera to further document the report. Finally, the Info screen gives a brief description of the application and the HealthMap system.



Figure 8. Client-side user interface for the Outbreaks Near Me application for iPhone. The Map and List views allow users to search and review real-time alerts. Users can submit a report through the Submit screen and enter their alert preferences on the Settings screen.

On the backend, the system is supported by a series of services running on the HealthMap server: the query service, the push service, and the report service. The query service accepts a range of queries and responds accordingly, with a list of alerts

organized by location. The most common query, by latitude, longitude pair, performs a reverse-geocoding* operation via Google’s Web service [36], and then applies the HealthMap location extraction engine to the resulting text, in order to determine the appropriate entity from the HealthMap location taxonomy. Once the location category is determined, all that remains is a Structured Query Language (SQL) query against the HealthMap database to return the appropriate list of alerts.

A search by location name is processed by first performing a geocoding step, again via the free Google Web service to determine the most likely coordinate match, followed by *reverse* geocoding and then all steps as for a coordinate-based query. The reason the initial geocoding step is required is that the HealthMap location extraction algorithm is designed specifically for processing location names from the text of news articles—it does not recognize key classes of place names, such as zip codes, although they are commonly entered by users for location queries, and readily recognized by the Google geocoder.

Another possible approach for these location-based queries is to use a distance calculation or a bounding box to identify all location entities within a given area. We chose not pursue this approach in part because of the computational overhead of the distance calculation, but primarily because of the way that we implemented the HealthMap location taxonomy. Specifically, for simplicity, each place entity has a single set of latitude, longitude coordinates associated to it, but the entity actually represents a spatial polygon, and the coordinates represent the approximate centroid of the polygon. Therefore, the distance between entities is really the distance between the centroids. Particularly for large polygons such as U.S. states, this distance thus becomes

* “Geocoding” refers to the process of resolving a place name to latitude and longitude; “reverse geocoding” consists of identifying a place name for a given coordinate pair.

problematic as a means for meeting the user expectation of seeing alerts in her local area.

Disease queries are significantly simpler to execute. The service applies the HealthMap disease name extraction engine to identify the appropriate category and performs the SQL query accordingly. Keyword searches are likewise simple to execute, as HealthMap employs a Lucene-based [37] keyword indexing system. Finally, in many cases, users may enter a query without selecting the appropriate search type, such as entering “h1n1” as a location, and the system is built with logic to infer the user intent and still process the query appropriately. In all cases, the service returns its results encoded in a standard JavaScript Object Notation (JSON) format for rendering by the client application.

Beyond the search service, the second key Web service supporting the application is the push engine. This service receives the values from the Settings screen and as alerts are collected by the HealthMap system, in turn sends out the appropriate alert messages to users. The first step, processing the user settings, consists of a similar geocoding and parsing procedure to determine the applicable location and disease entities, as for handling search queries. The push settings are then stored in the database as a record containing the user’s email address, unique iPhone device identifier, frequency preference, and the disease and location identifiers corresponding to the information submitted. The alert delivery service then runs hourly, querying the user preferences, and then querying the HealthMap database for the latest alerts matching each preference record, and either sending the alerts via email or delivering them to the phone via Apple’s “push” service.

Finally, the third backend component receives the user-submitted reports and saves them in the database for Web-based review by a moderator. It also handles storing

the image file if the report includes a photograph; the iPhone client uploads the image as coded text so it must be converted to binary for moderator review. The submission review and posting process is detailed below in Section 4.5.

Building on the work of the iPhone application, we then created an application for Android smartphones (Figure 9). The Android version works in much the same way as the iPhone application, although it is supported on a much broader range of phones and carriers, making it accessible to a broader range of users. In particular, the Android platform is increasingly becoming available on low-cost commodity handsets, making it suitable for use in developing countries, where many hotspots for emerging infection are located.

For the Android version, we are able to reuse the backend services from the iPhone app. The only subtle difference is that because Android phones support much higher resolution photography, when submitting an outbreak report, the Android application must upload the image via binary protocol. The text protocol used in the iPhone application doubles the size of the transfer payload, and for the large Android images, proved prohibitive under the bandwidth available on mobile networks.



Figure 9. *Outbreaks Near Me* for Android smartphones. As with the iPhone application, the user can view the latest alerts for her area as well as submit reports.

4.4 Voice and Web Applications

To further increase accessibility, we also created the HealthMap hotline, to allow users without smartphones, or even without cellular phones, to submit reports. Users can send a Short Message Service (SMS) message via cell phone to +1 919 MAP1-BUG (919-627-1284) or call and leave a voice message. The system then uses Google Voice [48] to convert the audio into text, which is in turn forwarded via email to the HealthMap system, and processed through the automated HealthMap input pipeline. Those who prefer email communication can also submit a report directly by emailing hotline@healthmap.org.

Finally, for the Web user, we also have two different Web-based community contribution services. First, we have a simple form appearing below the map on the HealthMap Web site, where the user can enter either a URL for a news article or a free-

form eyewitness report as noted above. Second, we have a Facebook application, currently in prototype form, that allows users to view a customized feed of HealthMap alerts on their Facebook page, share and comment on alerts within their social network, and, naturally, report events directly to HealthMap. The Facebook application is supported by still another Web service that supplies customized feeds of ongoing alerts based on user preferences. As with our mobile applications, the goals of the Facebook application are to generate attention for the project, provide an immediate benefit to the user, and serve as an easy and convenient reporting channel.

4.5 Clinical Reporting Tool

Last on our list of software tools is a new initiative we are calling “OutbreakMD,” whose impetus began with the January 12 earthquake in Port-au-Prince, Haiti. In the aftermath of the earthquake, Megan Coffee, M.D., Ph.D., an infectious disease physician volunteering at the general hospital (Hôpital de l'Université d'état d'Haïti, or HUEH) in Port-au-Prince, contacted Ushahidi seeking a system for disease tracking. Thanks to our ongoing collaboration, Ushahidi then forwarded the request to us and we have been working closely with Dr. Coffee to establish the system.

The HUEH is Haiti's largest hospital; despite lacking much of the infrastructure and supplies taken for granted in a modern hospital, it treats 169,000 patients in a typical year [38]. With the earthquake, the hospital was overwhelmed with the injured and dead [39]. While infectious disease was already a significant burden before the earthquake, its impact was exacerbated [40]. Meanwhile, Dr. Coffee is the only infectious disease physician currently working at the HUEH at the time of this writing in April 2010. The ability of her and her colleagues to track patterns in infectious disease is

particularly important for targeting interventions, both in bringing health care resources directly to internally displaced populations, as well as targeting environmental factors, such as spraying insecticides to prevent malaria or identifying and treating contaminated water sources.

The OutbreakMD system works as an “offline” Web application, taking advantage of the new HTML5 standard, supported in most modern Web browsers, including the iPhone and Android browsers, allowing dynamic Web pages to store data persistently in the browser. As shown in Figure 10, the application presents a simple form based around the key disease and symptom categories active in Haiti. When the user submits a report, the application saves it in the client-side SQLite [41] database and attempts to upload it to the HealthMap server. If no network connection is available or the connection is unreliable and it fails to upload, the report is stored for later upload. Once the user is online, she can synchronize the client with the server. To accomplish synchronization, we assign each report a globally unique identifier when it is created, and then transfer information in a three-phase handshake: first, the client submits the unique identifiers of its reports to the server, and the server compares them to its list. The server then returns both a list of the identifiers unknown to it (a subset of the client’s list) and full data for the reports the client does not yet have. Finally, the client stores the new reports from the server and uploads full data for the reports unknown to the server.

While this clinically oriented application constitutes a departure from the vision of engaging ordinary people as active participants in public health, it does exemplify a new model for reporting. Specifically, it overcomes many of the disadvantages of the traditional model of public health reporting by leveraging a more informal reporting channel for rapid information dissemination. Moreover, as noted below in Section 6.3.4, given that clinicians made use of our general-use iPhone application, there is a clear

need for this approach in the clinical setting. Although this particular application is still in the prototype phase, it has potential to have a significant impact in a place where help is sorely needed.

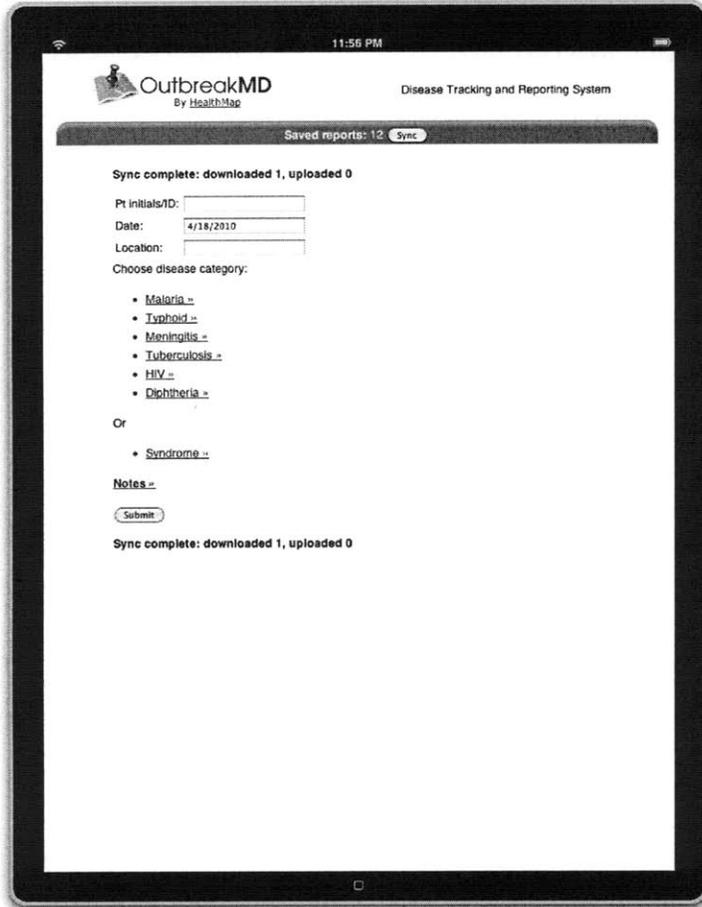


Figure 10. The OutbreakMD offline HTML5 application, shown here running on the Apple iPad, is intended to support clinicians to rapidly enter case reports either in the clinic setting or the field.

4.6 Submission Review and Validation Methods

As contributors use these various tools to submit information, each report enters an online queue for review and validation (Figure 11). The tool allows the reviewer to view the report in full, including the photograph, and the disease and location entities extracted by the automated parser. In many cases, the report can be approved or

dismissed by rapid review of the content. In cases where further investigation is needed, by clicking on the “Map” link, the moderator can immediately see the distance between the GPS coordinates reported by the phone and the location entered by the user. If the two points are close together, we can use that knowledge together with the content of the report to validate the submission to a certain degree. In some instances, especially if the report concerns a public event such as a school closure or a death, the moderator can consult additional sources (for example the school Web site or a local newspaper) to cross-validate the assertion. For the approval process, we employed a generally low threshold for acceptance: if a given report did not contain any indicators of obvious falsehood, it was approved. Some reports were rejected because they were of a disease that would be extremely unlikely to appear in the alleged location (such as African horse sickness in the U.S.), or the GPS coordinates were far from the reported location, and the report contained no additional indicators of veracity and was unable to be corroborated through news or other sources. The moderation tool also allows the reviewer to annotate a submission with comments, for example, to provide a link to a corroborating source or to indicate why a submission was approved or rejected in an ambiguous case.

Once a report is approved, it is geolocated to town-level resolution, any personally identifying information is removed, and it is posted directly to the map. The system also generates an email to the contributor thanking him for his submission, and inviting him to create a HealthMap user account and take credit for the submission.

Review User Submissions

With description only With photo only

Page 1

ID	1	Email	Coords	Date	Summary	Description	Diseases	Place Name	Places
2385				2010-04-23 09:14:21	Skin Rash	Three people with skin infections (boils, blisters, peeling skin); possibly parasitic infections	Rash	Ft. Worth, TX	Texas
2385		redacted		2010-04-22 10:43:05	Family with conjunctivitis	Kids got conjunctivitis from daycare and spread to parents.	Conjunctivitis	Manfield, MA	Massachusetts
2384		redacted	40 501, 75 883	2010-03-27 09:47:58					
2383		redacted	40 501, 75 883	2010-03-27 09:47:58					
2382		redacted		2010-03-27 09:47:58					
2381		redacted		2010-03-27 09:47:58					
2380				2010-03-27 09:47:58					
2379		redacted	40 501, 75 883	2010-03-27 09:47:58					
2378		m	40 501, 75 883	2010-03-27 09:47:58					

Back to list view

User Submission ID 2356

Summary: Fifth disease confirmed in Bronx NY

Description: School in Bronx confirms Fifth Disease - which is the last of the major diseases - after Mumps.

Email: redacted

Diseases:

Places: New York City

Coordinates: 40 501, 75 883

Place Name: redacted

Lat/Lon & Place Name: Map

Date: 2010-03-27 09:47:58

Report Status: Public

Alert ID: 402484

Supporting Alert ID:

Reviewed Status:

Medium: iPhone

Show full info

Edit corresponding alert

Add a Comment

Comments:

Emily Chan (Protected), 29 Mar 9:39: didn't see anything in Google News yet

Figure 11. Moderation queue for user submissions. Personally identifiable information such as email address, location coordinates, and physical address has been redacted in the figure and is not included in the public posting.

5. Community Response

Response to both the concept of community-based public health reporting and the specific software tools we developed has been very positive. Although we introduced the capability for users to submit URLs of news articles as a subtle change to the HealthMap Web interface, with little fanfare, our users immediately began entering valuable news articles. As described in detail below in Section 6.1, the feature was essential to our ability to track H1N1 infection and saw strong usage during the chaotic first wave. The concept has also, with a few exceptions, been well-received by public health professionals and officials more accustomed to traditional approaches.

5.1 Software Tools for the General Public

Our iPhone application, launched September 1, 2009 amid dozens of other H1N1-related applications from a range of institutions, was an immediate success (see Appendix A for a detailed review of other H1N1 applications for iPhone and Android). Immediately following its release, the application and the concepts behind it were featured in a broad range of news media outlets, including Wired News, the Wall Street Journal, the Boston Globe, CNN.com front page, Scientific American, Time, USA Today, New England Cable News, and the NBC Nightly News with Brian Williams, among others (<http://healthmap.org/press.php>). The application rapidly gained 100,000 downloads and was briefly the top-ranked free app in the Healthcare & Fitness section of the Apple iTunes store (see Appendix B) just as the second wave of H1N1 infection was beginning in the northern hemisphere. With the surge of users, we quickly exhausted Google's limit of 25,000 geocoding requests per Internet protocol address per day in the days following the launch. To work around the limit, we established an ad hoc network of

my own personally maintained servers to divide up the requests to Google and pass the responses back to the search algorithm. With the Android application, we saw less attention and usage, but it was featured in Verizon's online "store" in the Android Market and has been downloaded nearly 10,000 times. At the time the applications were released, iPhone market share in the U.S. was nearly seven times that of Android [47], and approximately 10% of iPhone Outbreaks Near Me downloads were to iPod devices not counted in smartphone market share statistics. With these adjustments, the Android usage is actually on par with the iPhone usage, for the relative level of Android adoption.

Although the HealthMap telephone hotline provides increased accessibility for reporting, very few contributors have used it, most likely because it doesn't provide any direct service to the user, in the way the applications and Web site provide a wealth of real-time information. Further, the phone number is a U.S.-based number, while the majority of U.S. users have either Web or smartphone access readily available, and therefore little reason to report by SMS or voice. We did, however, receive a report via the hotline of a cluster of unidentified skin infections occurring in Texas and originating in Mexico. We are currently investigating the report further; although it's impossible to make a diagnosis, it's possible that these are cases of cutaneous leishmaniasis, a parasitic infection endemic in Mexico but rarely seen in the U.S., and an important event for local health officials and clinical providers to be aware of.

With our most recent applications, we have successfully tested OutbreakMD with Dr. Coffee in Haiti, and are in the process of testing and deploying the Facebook application. While both appear poised to play important roles in the HealthMap reporting ecosystem, at this point it is still too early to gauge adoption or user response.

5.2 Response in the Professional Public Health Community

In October 2009, shortly after we released the iPhone application, I met with officials at the Boston Public Health Commission (BPHC) to demonstrate it to them and discuss possibilities for collaboration. While the officials did ultimately recognize some potential value for community-contributed public health reporting, particularly through aggregating submissions, they were generally against the approach. The reasons they outlined were that such a system would generate excessive noise and provide little added value without rigorous verification. In addition, they felt that in their role as government officials with a mandate to respond to citizen inquiries, such reports would only add to their burden without advancing their core mission. Finally, for their activities in early detection of disease, the BPHC makes use of a syndromic surveillance system that tracks emergency room visit data across Boston hospitals, and the officials felt that the system was already effective and sufficient for their needs. However, they did recognize that in certain cases, trends indicative of a significant event might emerge from a collection of individual reports, each of which on its own may not yield any worthwhile information.

By contrast, other public health officials I spoke with expressed enthusiasm about the community reporting concept and its potential to complement traditional public health structures to improve the system as a whole. At the International Congress on Infectious Diseases meeting in March 2010, I met with Aaron Kite-Powell, M.S., an official with the Florida Bureau of Epidemiology. He has been an active contributor to HealthMap, rapidly identifying outbreaks in Florida from the beginning of our efforts to engage users. While he felt that public health officials like him were still exploring how best to use community-contributed information, he said that it definitely held value and represented an important new direction for public health reporting. Similarly, Larry

Maddoff, M.D., an official with the Massachusetts Department of Health, underlined how important unofficial and unverified sources can be for early warning.

6. Results

The strong user response and the number of reports submitted showed that this new approach to public health reporting has potential for significant impact. By looking at the submissions collectively as well as individually, we uncovered important insights into how participants engaged with the tools and in turn how to use the system for early warning of disease threats.

6.1 Overview of Data Collected

We divided submissions into four basic types: ones referencing a news article; free-form “eyewitness” accounts of local events; personal accounts of illness, either of the submitter or close associates; and finally, reports that are not pertinent and to be hidden from view. In the pilot phase during the first wave of H1N1 infection, 88% of reports (all referencing news articles) were considered suitable for posting to the map. In the second phase, after implementation of mobile submission, across all submission channels, 74% were considered suitable with 20% news-based, 41% eyewitness, and 13% personal. Figure 12 illustrates how the distribution of reports across these categories varied over time during the study period, while Figure 13 shows a snapshot of the broad geographic distribution of reports. Out of all submissions, 5% included a photograph, and of these 27% were approved for posting. A sample of user-submitted photographs can be found in Appendix D.

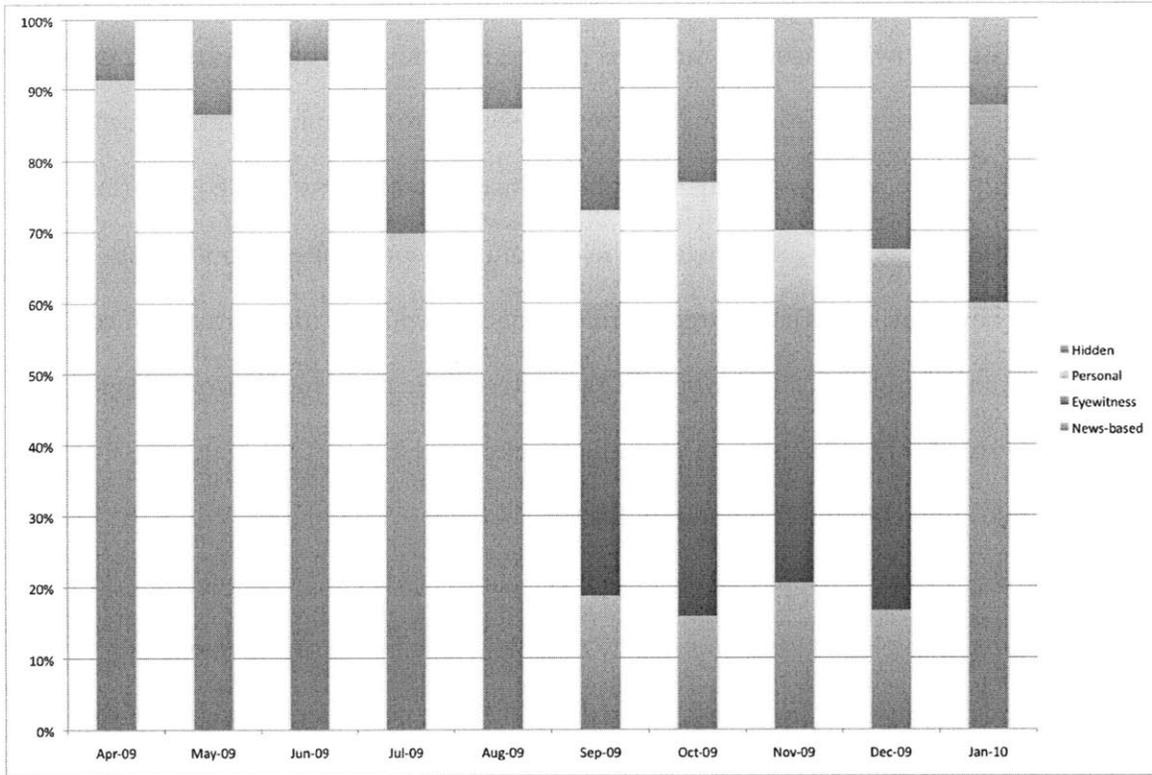


Figure 12. Distribution of user-submitted reports across the four categories: news-based, eyewitness, personal, and hidden. We introduced the ability to submit personal and eyewitness accounts on September 1, 2009.



Figure 13. Geographic distribution of events reported by HealthMap users. As most of the reports came from iPhone users, the distribution reflects the adoption of the iPhone, mainly by those in wealthy countries.

The pilot phase of the system coincided with the initial wave of H1N1 infections in April 2009, during which user submissions proved invaluable in our efforts to track the spread of the disease and provide an informative service to the public. In the months of April and May 2009, we saw 530 article submissions, of which 437 related to influenza, covering 40 countries. The advantage of the community-contributed reports was immediately evident: users submitted a broad range of reports from sources not covered by our automated systems, such as school Web sites or the news media in languages not available in HealthMap. As just one example, we received a report from an astute user of an H1N1 case at an elementary school in East Wenatchee, Washington. While the report is of just a single case, it is notable because it is a confirmed case, announced by an official source (the school), and thus highly reliable, while at the same time, the information is highly local in nature, and exactly the type of weak early signal of disease activity that is difficult to detect through news-based monitoring or traditional public health reporting. In another example, during early proliferation of H1N1 cases, on April 29, 2009, a user submitted a local Lithuanian-language news report of the first diagnosis to appear in Lithuania. While our automated news collection engine loaded the event one day later in English and other languages, thanks to an active HealthMap participant, we were able to detect the event in a timely manner and ahead of most other online monitoring tools.

Further, as noted above in Section 2.2, at times during the crush of news reporting of this period, the existing HealthMap content acquisition engine was overwhelmed with the sheer volume of information, making it difficult to identify the key reports. These user-submitted Web-based reports thus filled important gaps and improved the service for hundreds of thousands of site visitors.

The second-phase participatory epidemiology capability launched on September 1, just as the second wave of H1N1 infections was beginning in North America. As of April 30, 2010, the iPhone application had been downloaded 107,429 times, and had collected nearly 2400 submissions from users around the world. The majority of these (69%) were approved for publication, and of these, 95% pertained to influenza. Although corroboration or verification of individual reports is often difficult, we were generally able to filter spam, duplicates, and mistakenly submitted reports.

We took two approaches to evaluating these submissions. First, we constructed a small sample of reports with clearly identifiable events and determined whether the user-submitted information was novel or timely as compared to data from existing HealthMap sources, as shown in Table 1. With outbreaks at schools in particular, in several cases individual contributors provided timely information that was unavailable through other channels. In general, although it was difficult to pursue all possible sources of corroborating information in every case, for approximately 7.5% of reports (148 out of 1,987), we were able to locate some form of supporting news article or other information from a third party source.

Date	Category	Location	Excerpt	Info from pre-existing HealthMap sources	Comment
10/20/09	Eyewitness	Greendale, Wisc.	Canterbury elementary school closed until 10/23 due to 30% percent of students out with flu.	None.	This information was also on the school's official Web site.
10/28/09	Personal	Martinsburg, W. Va.	First my 5 year old son got it then my 18 month old daughter got it. Now my wife and I both have it.	None.	The submission also includes a report of crowding in the pediatric clinic.
9/5/09	Personal	Lafayette, La.	16 year old male with undocumented H1N1 with onset 1 week ago. Symptoms of fever to 104 F, sore throat, body aches and nonproductive cough for approximately 4 days with Relenza treatment.	None.	Detailed report, most likely coming from a clinician.
9/14/09	Eyewitness	Candia, NH	3year old infected with EEE	Confirmed in press reports later the same day.	Not all reports were of H1N1. This report was the first EEE case of the New England season.
4/23/09	News-based	Mexico	Canadians returning from Mexico urged to be on alert for flu-like symptoms.	Many media reports immediately following the user submission.	Many media reports of the event were collected through the existing HealthMap system; this submission was among the first indicating international spread.
9/17/09	Eyewitness	Charleston, SC	3 cadets have been placed into quarantine but since then 2 have returned with no more symptoms	None.	An H1N1 outbreak at the Citadel Military College was later reported on the school's emergency information Web site.

Table 1. Selected user submissions and the relationship to information from other HealthMap data collection channels.

As a second analysis, we aggregated the submissions and compared their fluctuations in volume over time with the variation in Centers for Disease Control and Prevention (CDC) influenza-like-illness (ILI) metrics for sentinel clinical sites in the United States [42]. Because the vast majority of community reports came from users of

the iPhone smartphone application (92%), for simplicity, we limited the aggregation to iPhone-originated reports only. We then computed the number of H1N1 submissions per application download, to mitigate bias from increased application usage driving increased reporting. Although the result is preliminary and not broadly generalizable given the special circumstances of H1N1, our aggregated metric correlated highly with ILI metrics (Pearson's correlation = 0.74, $p < 0.0001$, 95% confidence interval [0.44, 1.0]) as shown in Figure 14. There is a second peak in our activity metric occurring in late December 2009, where the CDC data indicate only a very modest increase in activity. One possible explanation for the divergence is simply that at that point in the epidemic, with incidence dropping and the public's attention shifting away from H1N1, the sample size of submitted reports became too small to serve as a strong indicator. However, another possibility that merits further investigation going forward is that the divergence may be due to the differing nature of the two indicators. While the CDC metric tracks clinical visits, the user submissions are often personal accounts of illness, relatively mild cases, where those infected may never seek clinical care. Without any interaction with a clinical provider, these cases will not appear in the CDC influenza-like-illness metric.

Overall, the strong correlation suggests potential use for aggregating user submissions as an additional early indicator of flu activity. Notably, while CDC flu metrics are generally released at least one to two weeks after they are collected, our metric can be made publicly available in near real-time as reports flow into the system. Further, the cost and difficulty of achieving buy-in from health care provider networks and establishing the nationwide ILI monitoring network is significantly greater than the resources required to create and maintain the HealthMap Outbreaks Near Me system.



Figure 14. Adjusted volume of H1N1 reports from HealthMap users (blue dashed line) as compared to weighted influenza-like-illness visit data from the CDC (red solid line), from September 5, 2009 to January 30, 2010. The Pearson’s correlation between the two values is 0.74.

6.2 Spam and Otherwise Unhelpful Reports

With the first-phase system whereby we invited users to submit Web content by pasting a URL on the HealthMap site, we have seen very little “spam,” or otherwise unwanted postings—as indicated above, approximately 12% of submissions were hidden. Of those, a few were examples of people simply advertising an unrelated site, either with the intention of having HealthMap users click on the link, or possibly with the hope of taking advantage of HealthMap’s Google search rank to improve their own rank. Meanwhile, the yield of the free-form submission channel (for example, the iPhone and Android applications) is significantly lower, with only 69% of reports accepted. The most likely explanation is that although these reports don’t go straight to the map, there is no requirement to identify an existing Web site—the user can simply type in a few letters or words and tap submit. For most of these rejected submissions, it is difficult to say exactly why they were submitted, as there is no economic incentive as there might be with posted links. The best explanation is either curiosity to see what will happen, or the simple urge to vandalize. In our system, reports are rejected silently, without any feedback to the user, in order to limit any chance for provocation.

6.3 Analysis of Selected Sample Submissions

Because the prompts and structures for user input are relatively open-ended, in order to reduce the burden on the user as noted above, we received a large diversity of reports. Given the broad variation in timeliness and significance across the submitted content, we can gain insight into the potential for the approach by examining and analyzing a collection of notable submissions.

6.3.1 News-based Submission via Mobile

With the smartphone applications, the submission form allows the user to attach a photograph to the report. While we weren't entirely sure how contributors would use the photo option, we had envisioned them using it to provide supporting evidence for their reports: pictures of dead animals or crowded medical facilities, for example. While many of the submitted pictures were jokes or otherwise unrelated to the report, some alert users helpfully photographed news articles describing outbreak events. Two such examples are shown in Figure 15, one in Swedish, the other in French. In the case of the Swedish article, the submitter also provided a translation as part of the report. While this type of report was rare, it shows the potential value in our approach, as in this case we successfully crowdsourced the discovery of an important event in a local language not implemented in HealthMap, as well as the translation of the content into English.

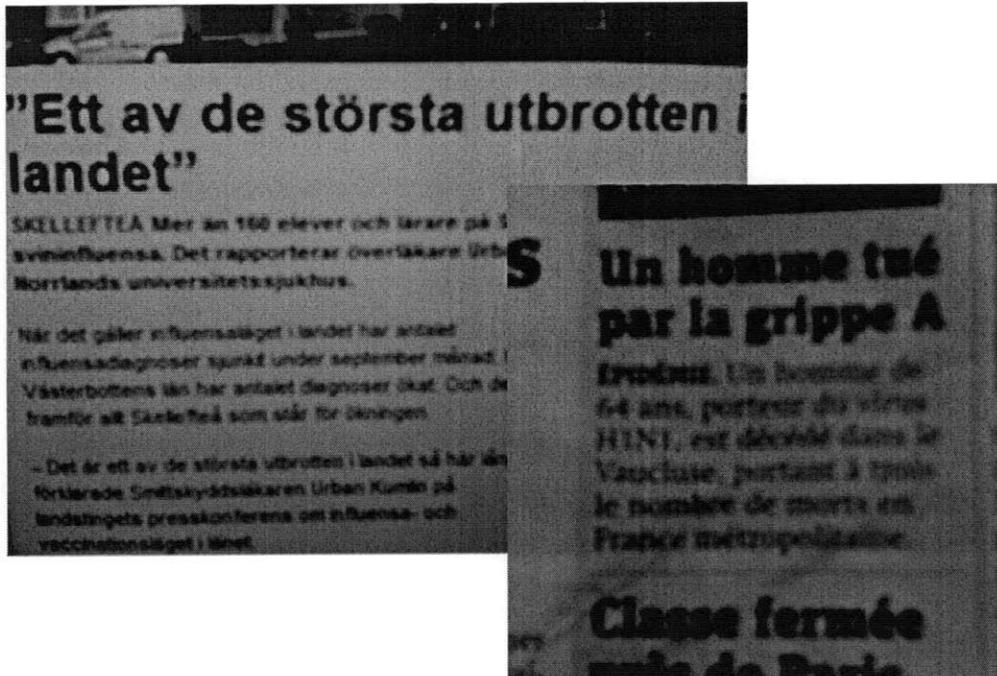


Figure 15. User-submitted news articles captured with the iPhone camera. Although we received few reports of this form, the technique served as an innovative way of documenting a submission. The contributor generally summarized the content in the text of the report as well.

6.3.2 Local Clusters and Personal Accounts

Although it remains difficult to verify or determine public health significance of individual reports of illness, these types of submissions hold promise because we currently have very limited means of accessing this information through existing channels. Examples of this type of report include (1) “First my 5 year old son got it [Influenza A H1N1] then my 18 month old daughter got it. Now my wife and I both have it. When we took our kids to the pediatrician's office it looked like a war zone. The doctor told us she'd been seeing them in wave after wave.” (2) “i had hands feet and mouth disease [sic],” and (3) “2 people. my husband - possibly from eating raw oysters. myself from contact with my husband. clearly norovirus symptoms. vomiting, headache, fever for 2 days.” In example 1, most likely none of the reported cases led to hospitalization, but it does demonstrate at a local level how rapidly H1N1 spread in a

largely immunologically naïve population, and the resulting clinical crowding. Example 2 illustrates the challenge of interpreting many of the submissions, both on account of users' lack of infectious disease expertise and the limitations of the iPhone's touch screen keyboard. In this case the user is reporting a case of hand, foot and mouth disease, generally caused by coxsackieviruses and frequently affecting infants and children. While complications are rare, infection can spread in daycare or school settings, and in certain cases cause deadly encephalitis [43]. Finally, the third example is of potential interest for public health authorities as it pertains to a contaminated food source as well as a contagious disease with high morbidity. Part of future research is finding ways of identifying these key reports consistently as well as determining appropriate alerting and response action measures.

6.3.3 Closures and School-Based Outbreaks

Many of the submissions were reports of closures of facilities due to widespread flu infection. Examples include reports such as “Outbreak of laboratory-confirmed H1N1 in schools in the provinces of Lucca and Pisa, Tuscany, Central Italy,” “Canterbury elementary school closed until 10/23 due to 30% percent of students out with flu,” and “Brown's Valley elementary school closed after 6 yr old girl died Sunday from confirmed H1N1 flu.” Notably, in the case of the Canterbury schools, we were able to verify the accuracy of the latter report as the information was posted on the school's public Web site. In the case of the Brown's Valley school, the report was submitted by a local (based on phone GPS coordinates) resident, and we were able to verify the death in the San Francisco Chronicle [44]. In all, we received close to 300 reports (approximately 10% of total reports) relating to outbreaks in school settings over the course of the study period.

6.3.4 Clinical Reports

Although the system is designed for laypeople, we also received a small number of reports appearing to come from clinical workers. A few such examples: (1) “16 year old male with undocumented H1N1 with onset 1 week ago. Symptoms of fever to 104 F, sore throat, body aches and nonproductive cough for approximately 4 days with Relenza treatment. Now afebrile for 2 days, but admitted to hospital for acute appendicitis”; (2) “09/05/09 Influenza A confirmed per nasal swab test. Started on Tamaflu [sic]. 15 year old male with history of asthma.” While we can’t confirm that the reports come from certified clinicians, the details and terminology reflect common clinical practices. They are therefore most likely truthful, and more importantly, show that clinicians also see value in our approach despite existing infrastructure that connects providers with public health authorities. In particular, these reports point to the promise of our new OutbreakMD system for facilitating direct clinical reporting.

7. Discussion and Limitations

Despite the potential for participatory epidemiology, many limitations remain. For one, particularly for endemic diseases, sickness in the population is a normal part of everyday life. However, user contributed data lack ongoing background rates of disease and negative case information that helps determine whether a given report represents a notable anomaly, or an ordinary occurrence. A related issue is how to motivate more people to report, especially on personal or local events: while a diligent searcher can generally find a steady volume of outbreak news from sources around the world, if one happens to be generally healthy and live in a healthy community, there may simply be nothing worthwhile to report. While we want to enable people to submit as easily as possible, given our limited ability to verify local information, we must take precautions not to motivate people to submit false reports.

One of the key motivators for engaging citizens as reporters is to access better information from developing countries around the world, areas that are often underserved not only by traditional public health infrastructure, but also by the online media sources that are vital to the established HealthMap model. Meanwhile, these countries are important focus areas for detection of emerging threats, not only because the burden of infectious disease is greater and health infrastructure is less well-developed, but also because local climate and ecological factors in tropical countries appear to foster novel pathogens more readily [6]. While we were able to achieve significant user engagement by leveraging the power and popularity of the iPhone platform, Figure 14 illustrates how the geographic distribution of the reports we received mirrors the distribution of iPhone users. As such, it tends to favor wealthier countries. As noted above, we have attempted to address this issue by providing, for example, an SMS-based reporting channel. However, because we have thus far struggled to provide this

user population with ready access to the existing HealthMap system and its wealth of real-time information in a way that is compelling and relevant to users' daily needs, it is unrealistic to expect contributions. Participatory epidemiology works best through a vibrant bidirectional flow of information.

Perhaps the most significant concern with our approach is the question of how to corroborate or verify submitted information. Public health officials may rightfully have reservations about this type of data: as observed above in Section 5.2, their obligation to respond to individual reports could represent an added burden to their surveillance responsibilities. However, we demonstrated some preliminary ways of analyzing the data through cross-validation with other sources (Table 1). Further, we envision the system as a two-way information exchange. Rather than simply supplying the government with reports, we publish the information on the HealthMap site so that community members can review and assess submissions. We also collect contact information from the person reporting, so in certain cases we have contacted the submitter to request additional details if a report raises particular interest.

Finally, although the emergence of H1N1 yielded a unique and powerful historic opportunity to carry out this project and engage the public, another key question is to what extent the observed phenomena will translate to other novel epidemics, or even the normal seasonal ebb and flow of infection. The awareness of the general public was greatly raised through the appearance of H1N1, but now that it has faded from the headlines, it remains to be seen whether we can continue to succeed in engaging the public around infectious disease reporting.

8. Conclusions and Future Work

Overall, the tools and approaches developed over the course of the thesis work are still in the beginning phases but continue to run and are actively used in production, and show tremendous potential going forward. In addition to maintaining and improving the breadth of our existing technologies, we are also exploring a number of key areas for new work.

In particular, we are working to open the process of reviewing and curating both news and user-contributed content, to allow more people to aid in the effort to effectively filter and characterize the large volumes of information we collect. To this end, one new feature we will be releasing on the Web site in the next few weeks is the ability for visitors to quickly flag a given report as “interesting” or “uninteresting” via a Digg-style [45] “thumbs up/thumbs down” button. With this type of user-engagement, we hope to rapidly determine events of interest, and further engage users who may have a valuable opinion on ongoing events in the world but no particular events to report from their own knowledge.

Outside the HealthMap site itself, Internet users are increasingly using “microblogging” services such as Facebook and Twitter to provide real-time updates on their personal status, including their illness and health. While initial efforts by others [46] have thus far found it difficult to extract meaningful signal from the vast volume of information, as we build on the thesis work, we plan to explore ways of tapping these services.

Another promising area of research, particularly for mobile users, is to examine passive collective intelligence through “reality mining” approaches, where users agree to be tracked via their phones, and we can then map and aggregate community interaction

patterns and how they relate to the spread of disease and its reporting. As observed above, travel clearly plays a key role in infectious disease transmission; with a large enough volume of consenting users, a better understanding of travel networks and flow volume would likely yield interesting insights for prevention, detection, and response measures.

Finally, with the knowledge gained from our work on infectious disease, we can explore how to apply participatory models and community-contributed information to other areas of public health. For example, we could potentially tap users to provide their experiences and insights on consumer product safety, or adverse events relating to pharmaceuticals and medical devices.

As is common with new approaches to existing challenges, especially approaches that seek to place more power in the hands of the untrained, unpredictable, and unreliable citizen, participatory epidemiology raises a degree of suspicion and skepticism from established experts. However, as we have demonstrated over the past year, with the appropriate tools and structures, there is clear value to be derived by engaging the general population as a highly distributed network of eyes on the ground, working together loosely but with common purpose.

9. References

1. Freifeld CC, Mandl KD, Reis BY, Brownstein JS. HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. *J Am Med Inform Assoc*. 2008 Mar-Apr;15(2):150-7. Epub 2007 Dec 20. PubMed PMID: 18096908; PubMed Central PMCID: PMC2274789.
2. Wikimedia Foundation. Wikipedia. <http://wikipedia.org>. Accessed 28 April 2010.
3. United States Centers for Disease Control and Prevention, Nationally Notifiable Infectious Conditions. <http://www.cdc.gov/ncphi/diss/nndss/phs/infdis2010.htm>. 2010.
4. Reingold, AL. Outbreak investigations—a perspective. *Emerg Infect Dis* 1998 Jan–Mar; 4(1): 21–27.
5. Keller M, Blench M, Tolentino H, Freifeld CC, Mandl KD, Mawudeku A, Eysenbach G, Brownstein JS. Use of unstructured event-based reports for global infectious disease surveillance. *Emerg Infect Dis*. 2009 May;15(5):689-95. Review. PubMed PMID: 19402953; PubMed Central PMCID: PMC2687026.
6. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P. Global trends in emerging infectious diseases. *Nature*. 2008 Feb 21;451(7181):990-3. PubMed PMID: 18288193.
7. Lacey, M. From Édgar, 5, Coughs Heard Round the World. *New York Times*. 28 April 2009. <http://www.nytimes.com/2009/04/29/world/americas/29mexico.html>. Accessed 28 April 2010.
8. Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD, Griffin J, Baggaley RF, Jenkins HE, Lyons EJ, Jombart T, Hinsley WR, Grassly NC, Balloux F, Ghani AC, Ferguson NM, Rambaut A, Pybus OG, Lopez-Gatell H, Alpuche-Aranda CM, Chapela IB, Zavala EP, Guevara DM, Checchi F, Garcia E, Hugonnet S, Roth C; WHO Rapid Pandemic Assessment Collaboration. Pandemic potential of a strain of influenza A (H1N1): early findings. *Science*. 2009 Jun 19;324(5934):1557-61. Epub 2009 May 11. PubMed PMID: 19433588.
9. Morales, A. Veracruz: reporta agente municipal extraño brote epidémico que ha cobrado dos vidas. *La Jornada*. 1 April 2009.
10. U.S. Centers for Disease Control and Prevention. Swine Influenza A (H1N1) Infection in Two Children -- Southern California, March-April 2009. *Morbidity and Mortality Weekly Report*. 21 April 2009. <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5815a5.htm>. Accessed 28 April, 2010.
11. World Health Organization. Influenza-like illness in the United States and Mexico. *Global Alert and Response, Disease Outbreak News*. 24 April 2009.

- http://www.who.int/csr/don/2009_04_24/en/index.html. Accessed 28 April 2010.
12. MSNBC. Swine flu events around the world. http://www.msnbc.msn.com/id/30624302/ns/health-cold_and_flu. Accessed 28 April 2010.
 13. World Health Organization. Current WHO phase of pandemic alert. http://www.who.int/csr/disease/avian_influenza/phase/en/index.html. Accessed 28 April 2010.
 14. New England Journal of Medicine, Journal Watch. H1N1 Influenza Center. <http://h1n1.nejm.org/>. Accessed 28 April 2010.
 15. Madrigal, A. iPhone App Finds Disease Outbreaks Near You. Wired Science. 1 September 2009. <http://www.wired.com/wiredscience/2009/09/outbreaksnearme/>. Accessed 28 April 2010.
 16. LaVallee, A. App Watch: A Mobile Swine-Flu Tracker. Wall Street Journal, Digits blog. 5 October 2009. <http://blogs.wsj.com/digits/2009/10/05/app-watch-a-mobile-swine-flu-tracker/>. Accessed 28 April 2010.
 17. Bazell, R. Tracking swine flu? There's an app for that. NBC Nightly News with Brian Williams. 16 September 2009.
 18. CNN. First doses of H1N1 flu vaccine arrive. CNN.com. 5 October 2009. <http://us.cnn.com/2009/HEALTH/10/05/h1n1.vaccine.campaign/index.html>. Accessed 28 April 2010. Collected via HealthMap.org.
 19. Malone TW, Laubacher R, Dellarocas C. Harnessing Crowds: Mapping the Genome of Collective Intelligence. MIT Sloan School of Management, Research Paper Series. 3 February 2009.
 20. InnoCentive. <http://www.innocentive.com/>. Accessed 28 April 2010.
 21. Threadless. <http://www.threadless.com/>. Accessed 28 April 2010.
 22. The ESP Game. <http://www.gwap.com/gwap/gamesPreview/espgame/>. Accessed 28 April 2010.
 23. Linden G, Smith B, York J. Amazon.com Recommendations: Item-to-Item Collaborative Filtering. IEEE Internet Computing. January-February 2003.
 24. Shardanand U, Maes P. Social Information Filtering: Algorithms for Automating "Word of Mouth." Computer-Human Interaction (CHI) Conference Proceedings. 1995.
 25. Page L, Brin S, Motwani R, Winograd T. The pagerank citation ranking: Bringing order to the web. 1998. <http://www-db.stanford.edu/~backrub/pageranksub.ps>. Accessed 28 April 2010.

26. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature*. 2009 Feb 19;457(7232):1012-4. PubMed PMID: 19020500.
27. Google Flu Trends. <http://www.google.org/flutrends/>. Accessed 28 April 2010.
28. International Society for Disease Surveillance. Distribute. <http://www.isdsdistribute.org>. Accessed 28 April 2010.
29. Reis BY, Kirby C, Hadden LE, Olson K, McMurry AJ, Daniel JB, Mandl KD. AEGIS: a robust and scalable real-time public health surveillance system. *J Am Med Inform Assoc*. 2007 Sep-Oct;14(5):581-8. Epub 2007 Jun 28. PubMed PMID: 17600100; PubMed Central PMCID: PMC1975800.
30. Madoff LC, Woodall JP. The internet and the global monitoring of emerging diseases: lessons from the first 10 years of ProMED-mail. *Arch Med Res*. 2005 Nov-Dec;36(6):724-30.
31. ProMED Mail. Pneumonia - China (Guangdong): RFI. ProMED Mail 2003; 10 Feb: 20030210.0357. <http://promedmail.org/pls/apex/f?p=2400:1001:4384177365297598>. Accessed 28 April 2010.
32. Ushahidi. <http://ushahidi.com/>. Accessed 28 April 2010.
33. Meier P. Haiti: Taking Stock of How We Are Doing. 6 February 2010. Ushahidi Blog. <http://blog.ushahidi.com/index.php/2010/02/06/ushahidi-how-we-are-doing/>. Accessed 28 April 2010.
34. Feldmann V. Mobile overtakes Internet: Implications for Policy and Regulation. *International Telecommunications Union* 2003:1-39. http://www.itu.int/osg/spu/ni/mobileovertakes/Resources/Mobileovertakes_Paper.pdf. Accessed 28 April 2010.
35. The Pew Global Attitudes Project. World Publics Welcome Global Trade—But Not Immigration. 4 October 2007. Page 7. <http://pewglobal.org/reports/display.php?ReportID=258>. Accessed 28 April 2010.
36. Google. The Google Geocoding Web Service. <http://code.google.com/apis/maps/documentation/geocoding/>. Accessed 28 April 2010.
37. The Apache Software Foundation. Apache Lucene. <http://lucene.apache.org/java/docs/>. Accessed 28 April 2010.
38. International Training and Education Center for Health. Spotlight: Clinical Mentoring at HUEH. <http://www.go2itech.org/where-we-work/haiti>. Accessed 28 April 2010.

39. Charles J, Clark L, Robles F, Daniel T. Death toll and injured overwhelms hospital in Haiti; damage stymies relief efforts. Miami Herald. 14 January 2010. <http://www.miamiherald.com/2010/01/14/1425787/death-toll-and-injured-overwhelms.html>. Accessed 28 April 2010.
40. World Health Organization. Public health risk assessment and interventions: Earthquake: Haiti. January 2010. http://www.who.int/diseasecontrol_emergencies/publications/who_hse_gar_dce_2010_1/en/index.html. Accessed 28 April 2010.
41. SQLite. <http://www.sqlite.org/>. Accessed 28 April 2010.
42. Centers for Disease Control and Prevention (2009) US Outpatient Influenza-like Illness Surveillance Network (ILINet). <http://www.cdc.gov/flu/weekly/fluactivity.htm#OIS>. Accessed 28 April 2010.
43. British Broadcasting Corporation. China on alert as virus spreads. 3 May 2008. <http://news.bbc.co.uk/2/hi/asia-pacific/7381741.stm>. Accessed 28 April 2010.
44. Gordon R. Girl with swine flu dies in Vacaville. San Francisco Chronicle. 23 October 2010. <http://www.sfgate.com/cgi-bin/article.cgi?f=/c/a/2009/10/22/MNJU1A9KPT.DTL&tsp=1>. Accessed 28 April 2010.
45. Digg.com. <http://digg.com>. Accessed 28 April 2010.
46. DIYCity/SickCity. <http://sickcity.org/>. Accessed 28 April 2010.
47. comScore. comScore Reports February 2010 U.S. Mobile Subscriber Market Share. 5 April 2010. http://www.comscore.com/Press_Events/Press_Releases/2010/4/comScore_Reports_February_2010_U.S._Mobile_Subscriber_Market_Share. Accessed 30 April 2010.
48. Google. Google Voice. <http://www.google.com/googlevoice/about.html>. Accessed 3 May 2010.

Acknowledgements

This thesis work, truly a collective effort, was made possible thanks to the guidance, contributions, and support of many others. I first thank all of the many HealthMap users who participated in the community and contributed their knowledge and experiences to improve the system. Second, I extend warm thanks to my thesis readers and advisors, Frank Moss, Pattie Maes, John Brownstein, and Chris Schmandt, for their advice and encouragement. Thirdly, my thanks are due to the HealthMap team at Children's Hospital Boston, including Sue Aman, Sumiko Mearu, Amy Sonricker, and Emily Chan for their help in the development and testing of the tools as well as with the data collection. Special thanks also to Ben Reis for his collaboration on the HealthMap Facebook application. Finally, I extend my profound gratitude to my wife Andee for her enduring and unwavering support for my pursuit of graduate education.

Appendices

Appendix A. Review of Selected H1N1 Applications for iPhone and Android.

Swine Flu Smartphone Application Roundup:

A review of selected iPhone and Android applications for staying informed about H1N1

Clark Freifeld
December 15, 2009

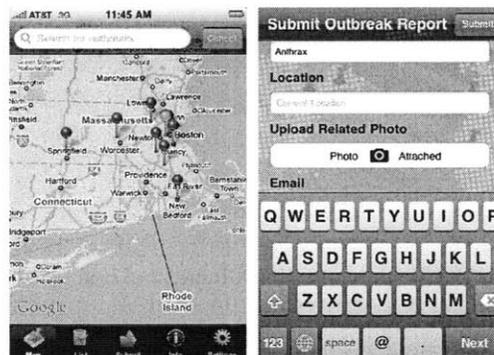
Introduction

Smartphone platforms such as iPhone and Android are the most rapidly growing personal computing systems currently available. As a departure from both traditional desktop PC interfaces and Web interfaces, phones represent an exciting area of human-computer interaction study: hardware and software are continually changing, novice users are adopting the devices in large numbers, and few conventions are established.

In parallel with the ongoing adoption of smartphones, the appearance of novel H1N1, or “swine flu,” in spring 2009 represents the first major emerging disease pandemic in recent years. As a result, developers around the world have created myriad interactive Web sites and software, including smartphone applications, based around H1N1 tracking. As a case study of smartphone user experience considerations, in this paper, we examine a selection of “swine flu” applications for iPhone and Android, comparing and contrasting their functionality and user interface design.

In particular, we are interested in how they compare to “Outbreaks Near Me,” the disease tracking and reporting application we developed based on the HealthMap Web-based system. We begin with a description of Outbreaks Near Me, its available features and limitations, both on the iPhone and Android platforms. We subsequently present a selection of iPhone apps, followed by a selection of Android apps, providing description and qualitative assessment of the user experience based on my testing.

Outbreaks Near Me



Outbreaks Near Me, the HealthMap smartphone application, provides essentially two key features: first, it delivers near real-time disease outbreak alerts, based on the user’s current location or other search criteria; second, it allows the user to submit an outbreak

report into the system, including an optional photograph to use as needed to supplement the report. Regarding the first feature, the alerts are displayed on an interactive map, as well as being available for delivery directly to the user via email or “push” based on the user’s preferences. As to the second, with its availability, we introduce “participatory epidemiology” where we allow users to contribute their own knowledge and experiences to improve HealthMap and benefit the community as a whole. Notably lacking from the app is any form of static content pertaining to swine flu, such as symptoms, prevention practices, vaccine efficacy, treatment options, and so on. Our reasoning in not providing this information is that users can for the most part obtain it from official health agency Web sites at their leisure, whereas the value of a mobile app is in its ready availability at all times, both for receiving continually updated content and for submitting event information as it happens.

iPhone

We begin with the market leader and dominant player in the smartphone field, Apple’s iPhone, with 100,000 apps available* and approximately 10 million users in the US alone.** A search for “swine flu” on the iPhone App Store yields dozens of matches, including games, jokes and other software only tangentially related to H1N1 tracking. After winnowing out entertainment-related apps, we are left with a handful of interesting options. We will review “Flu Tracker, ” “Swine Flu Tracker,” and “SwineFlu,” available for free; “Swine Flu,” available for \$0.99; and HMSSwineFlu, available for \$1.99.

* Rocha, Jerry. The Droid: Is this the Smartphone Consumers are Looking For? NielsenWire Blog, 11/11/2009. <http://blog.nielsen.com/nielsenwire/consumer/the-droid-is-this-the-smartphone-consumers-are-looking-for/>. Accessed 12/14/2009 and

Elmer-DeWitt, Philip. Apple fact check: 50,000 iPhone apps? CNNMoney.com, 6/10/2009. <http://brainstormtech.blogs.fortune.cnn.com/2009/06/10/apple-fact-check-50000-iphone-apps/>. Accessed 12/14/2009.

** Foresman, Chris. AT&T reports smooth sailing on the iPhone front for Q1'09. Ars Technica, 4/22/2009. <http://arstechnica.com/apple/news/2009/04/att-reports-smooth-sailing-on-the-iphone-front-for-q109.ars>. Accessed 12/14/2009.

1. **Flu Tracker/The Swine Flu Tracker**, by Garafa, LLC. Free.



Flu Tracker is one of the more comprehensive and easy-to-use apps on the list. Although the map is the less commonly used Open Street Map platform (as opposed to Google Maps), it works seamlessly for the application. The controls along the top allow you to (1) center the map on your current location (from phone GPS), (2) toggle display of confirmed H1N1 case markers, (3) toggle display of H1N1 death markers, and (4) toggle display of suspected case markers. Tapping a marker brings up a screen with the information for that location as culled from official and news media sources. Tapping the link to the original source brings up the original Web page within the frame of the application, making it easy for the user to navigate back to the map when done reading. The Information screen displays links to information about swine flu in Wikipedia along with videos from the CDC.

However, the app suffers from a few key limitations. First, the data displayed on the map have not been updated since mid June, 2009. Therefore, the application is missing any recent data, including all second wave outbreaks. It appears that the points shown on the map are “baked in” to the application, rather than loaded dynamically from a Web service. Even if the developer updated the data frequently, the user would need to upgrade the application just to receive new data, an inconvenient operation. The application also doesn’t allow the user to subscribe to location-targeted alerts and be notified as information becomes available.

As regards user interface, the application has very high usability in my assessment. The one criticism I have is that to submit a report of an outbreak, you double tap on the map location for the report. I only discovered the feature by accident and had trouble reproducing it because I hadn’t realized I had double-tapped. Following the double tap and confirmation, the application launches the email application with the body of the email pre-populated with prompts for information. The email approach has the advantage of capturing the submitter’s email address automatically (without typos) but

makes it more difficult for the user to attach an image, and impossible for the user to choose the disease category from a list of options.

2. **Swine Flu Tracker/Swine Flu Tracker Map**, by IntuApps. Free.

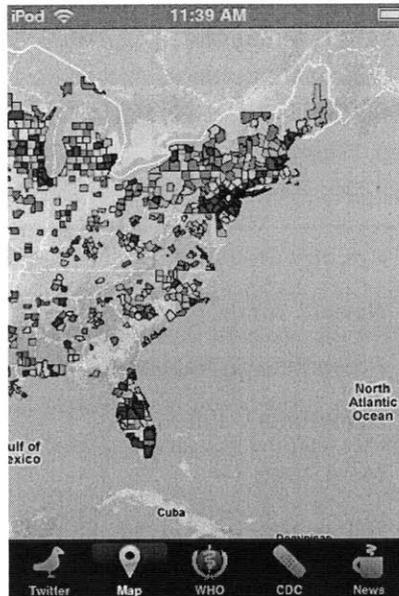


The Swine Flu Tracker app again uses Open Street Map data for its map display, although to my eye the aesthetics of the map in terms of color choice and layout are better than the map used in the Flu Tracker app. The map data appears to be relatively up-to-date (I found reports from late November 2009 on a cursory check), based on official and news media sources, although somewhat thin in coverage.

In terms of the user interface, the app opens first to the tab displaying the World Health Organization (WHO) official threat level (on the six-point scale). While the WHO designation is important, it doesn't change frequently, and, as a worldwide status, is necessarily less specifically relevant to me than local reports. As a user, likely my most frequent goal in opening the app would be to find out what latest activity is happening in my area. Under the design of this application, to accomplish the task, after launching the app, I have to tap the Map tab, then tap the Locate Me button. Further, while I can preview report information, when I tap a link for a full report, the browser application launches, and to return to the Swine Flu Tracker, I have to relaunch the app and navigate back to where I was previously.

Beyond the map, the app also includes an aggregation of RSS from the WHO, CDC and HealthMap Twitter and blog feeds, a useful one-stop-shop of relatively high-quality information. Another screen within the app displays static information on swine flu symptoms and background from the CDC Web site, although the information appears to precede recognition of the 2009 novel strain as it addresses primarily endemic seasonal swine influenza (historically affecting mainly pigs).

3. **SwineFlu/Swine Flu!** by Matthew C. Lawyer, MD JD MPH. Free.

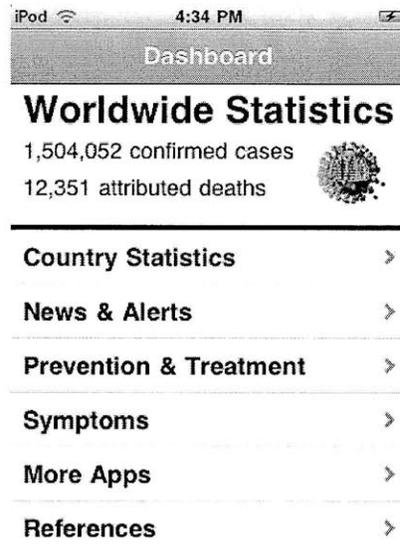


The last of the freely available apps examined for this review, “SwineFlu” brings together a number of different live RSS sources, including Twitter (#swineflu hashtag), WHO, CDC, and news media (search for “swine flu”). However, the information is difficult to filter and navigate as all four streams represent worldwide data, as opposed to being specifically relevant to my location or other criteria. Moreover, the Twitter feed includes significant amounts of noise in the form of jokes and other irrelevant information.

The app includes a map but it is not interactive; it is a screenshot of the Rhizalabs H1N1 tracking Web site taken on August 17th. From a user-interface standpoint, the map adds confusion because it is of the U.S. only and does not relate to or represent data from any of the other app screens.

One advantage the application does have over the others is that it works equally well in portrait as well as landscape mode, giving the user more viewing options. The above free apps work exclusively in portrait mode.

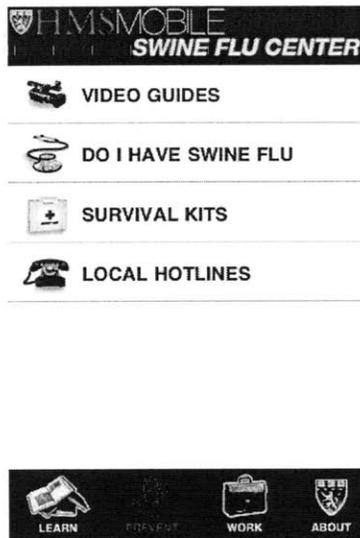
4. **Swine Flu**, by Keaka Jackson/DoctorCalc. \$0.99.



Although at \$1 this app is expensive given that it provides more limited functionality than some of the free apps, it actually has a certain appeal in that it is simple, clean, easy to use, presents validated information, and doesn't market itself beyond the reality of what it provides. Specifically, it presents an easily browsable and sortable list of H1N1 case and death data by country, continually updated from the WHO Web site. The app also includes a "News & Alerts" tab collecting RSS feeds from the WHO and CDC. The user interface for this view could be improved by labeling the items as to which of the two sources they come from; clicking an item launches the separate browser application, requiring the user to relaunch the Swine Flu app and navigate to the previous position upon completion.

The static content included with the app is very limited and not especially useful or beyond what one would find with some basic Web searches.

5. **HMSswineFlu/HMSMobile Swine Flu Center**, by Pocketized Ltd/Harvard University. \$1.99.



This app includes extensive static content, primarily in the form of videos featuring Harvard Medical School's Dr. Anthony Komaroff along with animations explaining flu mechanisms and disease spread. As a result of the large amount of video, the application bundle itself is 75 megabytes in size, resulting in slow installation, especially on a slow connection. As regards other static content, while a number of the other applications include listings of swine flu symptoms, the HMS Swine Flu Center differentiates itself by presenting an interactive symptom questionnaire entitled "DO I HAVE SWINE FLU." Based on the users answers, the system makes a recommendation as to whether the user may have swine flu and should seek help from a clinician.

For dynamic content, there is a map screen built into the app, based on the Outbreaks Near Me data service. It does provide up-to-date, location-based outbreak reports, but it's also available for free download with the Outbreaks Near Me application. On the Prevent screen, there is an option entitled News Feed, but at the time of this writing, it only shows three items from the CDC RSS feed.

Overall, while it has some dynamically updated elements, the core of the application represents a large trove of static content, mainly videos, pertaining to swine flu. However, my impression of the videos is that the ones starring Dr. Komaroff are not especially engaging to watch, while the animations, done in black, white and red with dramatic music and voiceover, seem to overemphasize the potential danger of the pandemic.

Android

While Android phones have to date significantly lower market share than the iPhone, Android is an important and rapidly evolving alternative in the smartphone space. Its advantages include being free and open source, allowing any handset manufacturer to build Android-compatible phones, as well as being a true multi-tasking OS, including a

framework for “widgets,” applications that appear on the phone’s home screen and run continually. The open nature of the platform means that it is available across service providers; in the U.S., Verizon, Sprint, and T-Mobile all currently offer Android phones, whereas the iPhone is available only to AT&T subscribers. This wider handset compatibility is also important for developing countries, where Android appears poised to make rapid inroads. The emerging market is particularly important for infectious disease monitoring applications, as it represents areas of the world that are disease “hotspots” where human-animal interaction is more frequent, and populations are more vulnerable.

Using a similar procedure as with the iPhone, I searched the Android Market for “swine flu,” garnering around a dozen matches. As before, many of these were jokes or games; I narrowed the selections to three: “Flu Tracker Free,” by Navee Technologies, “Swine Flu” by Eknath Kadam (also free), and “Swine Flu” by Androwave Inc, \$0.99. (Another notable advantage of Android over the iPhone is that I can return any purchase for a refund within 24 hours.)

1. **Flu Tracker Free**, by Navee Technologies. Free.



The app launches with a video tutorial, which starts out well but transitions to a bizarre robotic text-to-speech voiceover.

The main screen provides three buttons, Prevent, Track, and Symptoms. Tapping Track allows the user to select between an H1N1 map and a seasonal flu map. The maps, as depicted above, give case totals by country and, for the U.S., by state. There is no indication as to where the data come from and when they were last updated, but based on my experience with H1N1 data, they most likely come from the WHO Web site (as in

the tables presented in the “Swine Flu” app for iPhone by Keaka Jackson). No real-time, location-based outbreak event information appears to be available through the application. Beyond maps, there is a screen entitled “Alert,” which appears to show news articles and videos. However, all but one of the news articles current visible had no relevance to swine flu and the videos appear not to have been updated recently.

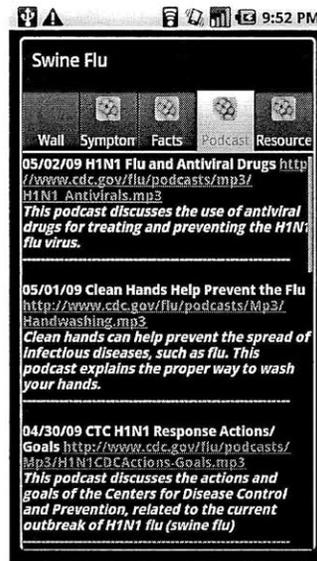
For static content, the app has a large complex of different screens with information on H1N1, seasonal flu, avian flu, canine flu (affecting only dogs), vaccines, multiple screens on daily prevention practices, and on. However, although the content appears authoritative, and contains clues that it comes from the CDC, there is no explicit indication of its source.

2. **Swine Flu**, by Eknath Kadam. Free.



The app contains only static content, basically consisting of an FAQ on H1N1, with questions such as “What is Swine Flu?” “Is there a vaccine available for Swine Flu?” and “How do I know if I have influenza A(H1N1)?” The information is meager and again not cited, making this app not worth the trouble to install.

3. **Swine Flu**, by Androwave, Inc. \$0.99.



The 99-cent Swine Flu application again represents little added value over the CDC Web site. It presents some basic information without attribution (again most likely taken from the CDC) and presents the CDC podcast feed, but the feed appears not to have been updated since May 2, 2009. For dynamic content, the app does offer updates on swine flu via SMS. However, judging by the other features of the app, I'm skeptical as to their value.

Conclusion

Based on this sampling of the available H1N1 smartphone applications, I can see two key conclusions. The first is not surprising but still interesting to see illustrated as starkly as it is in this case study: the Android platform lags significantly behind iPhone in both the number and quality of its available applications. Developers still see iPhone as the premiere smartphone platform and devote their resources first and foremost to creating competitive iPhone applications. From the user-interface perspective, the greater maturity of the iPhone platform is also evident from a quick look at the included screenshots: four out of five of the iPhone apps use the “row of navigation icons along the bottom” (in white-on-black) concept, which appears to be gaining traction as a convention for iPhone interaction. Meanwhile, the Android apps present a cacophony of different looks with varying degrees of aesthetics and legibility.

The second conclusion is that despite the number of lukewarm and negative reviews we have received for the Outbreaks Near Me application on both platforms, it is clearly among the leading smartphone applications for swine flu, certainly *the* highest quality application available for Android. On the iPhone platform, other applications have shown that key usability improvements, such as the ability to view news articles within the application, remain to be made to the Outbreaks Near Me application. However, especially in terms of providing dynamically updated information in an intuitive structure, our application provides significant, unique value to the user.

Appendix B. "Outbreaks Near Me" Is Top Free App in Healthcare & Fitness



Appendix C. End-user License Agreement for Submissions to HealthMap

By sending a written contribution to HealthMap, you grant a perpetual, royalty-free, unconditional license to HealthMap, and any successor organization, to publish your contribution on the HealthMap Web site and all affiliated sites, as well as disseminate it to other parties via RSS or other media, and to discuss or reference it in any publications related to or arising out of HealthMap. You also agree that HealthMap has the right, but not the obligation, to edit or remove any contribution, or to merge it in text or concept with other contributions, in HealthMap's sole discretion.

We will credit your contribution by publishing it with your name or an acceptable pseudonym of your choosing, or you may remain anonymous. We collect personal information such as, but not limited to, your email address and Internet Protocol address. We reserve the right to use this information internally, including for correspondence with the contributor, but will not sell, share, rent, or otherwise reveal this information to any third party except as required by law, or to address issues of noncompliance.

HealthMap and its users, including government agencies, depend on contributors for the accuracy of submissions. HealthMap takes no responsibility for errors or inaccuracies in any submission. Deliberate misrepresentation of information by a contributor may constitute a violation of law, and if it is serious will be reported to government authorities.

Appendix D. Selected User-Submitted Photographs.



