Social media in public health

Taha A. Kass-Hout* and **Hend Alhinnawi**

Humanitarian Tracker, Washington, DC, USA

Introduction or background: While social media interactions are currently not fully understood, as individual health behaviors and outcomes are shared online, social media offers an increasingly clear picture of the dynamics of these processes.

Sources of data: Social media is becoming an increasingly common platform among clinicians and public health officials to share information with the public, track or predict diseases.

Areas of agreement: Social media can be used for engaging the public and communicating key public health interventions, while providing an important tool for public health surveillance.

Areas of controversy: Social media has advantages over traditional public health surveillance, as well as limitations, such as poor specificity, that warrant additional study.

Growing points: Social media can provide timely, relevant and transparent information of public health importance; such as tracking or predicting the spread or severity of influenza, west nile virus or meningitis as they propagate in the community, and, in identifying disease outbreaks or clusters of chronic illnesses.

Areas timely for developing research: Further work is needed on social media as a valid data source for detecting or predicting diseases or conditions. Also, whether or not it is an effective tool for communicating key public health messages and engaging both, the general public and policy-makers.Keywords: digital disease detection, surveillance, communication, crowdsource, mobile, social media.

Introduction

Crowdsourcing, folksonomy, user-generated content, social networks, the sharing economy, peer production, Multi-User Virtual Environments, participatory media, collaborative creativity and Big Data (see Table 1) are all terms in the expanding lexicon of user-contributed data and user-led innovation (Sharp, Darren and Salomon, Mandy, User-led innovation: a new framework for co-creating business and social value, Smart Internet

*Correspondence address.
PO Box 2156, Arlington,
VA 22202, USA. E-mail:
kasshout@gmail.com*

Table 1 Important terms and definitions

Term	Definition
Crowdsourcing	Crowdsourcing is a neologism for the act of taking tasks by a group (crowd) of people or community in the form of an open call. The term has become popular for the trend of leveraging the mass collaboration. *Source*: Daren C. Brabham. 'Crowdsourcing as a Model for Problem Solving: An Introduction and Cases', Convergence: *The International Journal of Research into New Media Technologies* 2008;14(1):75–90
API	An Application Programming Interface (API) is a specification intended to be used as an interface by software components to communicate with each other. An API may include specifications for routines, data structures, object classes and variables. *Source*: Wikipedia http://en.wikipedia.org/wiki/Application_programming_interface
Folksonomy (or Tags)	A folksonomy is a system of classification derived from the practice and method of collaboratively creating and managing tags to annotate and categorize content; this practice is also known as collaborative tagging, social classification, social indexing and social tagging. Folksonomy, a term coined by Thomas Vander Wal, is a portmanteau of folks and taxonomy
	Folksonomies became popular on the Web around 2004 as part of social software applications, such as social bookmarking and photograph annotation. Tagging, which is one of the defining characteristics of Web 2.0 services, allows users to collectively classify and find information. Some websites include tag clouds as a way to visualize tags in a folksonomy. A good example of a social website that utilizes folksonomy is 43 Things
	An empirical analysis of the complex dynamics of tagging systems has shown that consensus around stable distributions and shared vocabularies does emerge, even in the absence of a central controlled vocabulary, published in 2007 (Halpin *et al*., 2007). For content to be searchable, it should be categorized and grouped. This is possible only if the content is tagged like keywords in a journal article. *Source*: Robu, V., Halpin, H. and Shepherd, H. Emergence of consensus and shared vocabularies in collaborative tagging systems, *ACM Transactions on the Web (TWEB)*, 2009;3(4), art. 14. Halpin, H., Robu, V. and Sheperd, H. 'The complex dynamics of collaborative tagging', In: 16th International World Wide Web Conference (WWW-2007). Banff, Canada: W3C, 2007, pp. 211–220
Geotagging	Geotagging (also written as GeoTagging) is the process of adding geographical identification information to various media such as a photograph or video, websites, SMS messages or news source and is a form of geospatial metadata. The data usually consists of latitude and longitude coordinates, though they can also include altitude, bearing, distance, accuracy data and place names. *Source*: Wikipedia http://en.wikipedia.org/wiki/Geotagging

Technology CRC; 2008 http://hdl.handle.net/1959.3/23016). Central to each term is a common concept—the co-operative active user.

Internet users can communicate online via tools that allow social interactions; hence Social Media (hanging out with friends, making new friends, playing games, sharing insights and stories, 'researching' the Internet). The Internet is a novel source of contextualized health data that has enabled an explosion of user-driven innovations that monitor infectious and chronic disease trends in populations. Every minute, people around the world are publicly sharing large volumes of personal information about their health and the health of their communities. Timely interactions on social media also permit reciprocity in how people share information online, the more one shares, the more likely others will share back. Some are using brief 140

character statements on Twitter, Facebook, Google+ and other social networking platform. Some are writing lengthier expositions about their health in discussion forums or blogs, and everyone is leaving a trail of health questions on Internet search engines. Social data are made easily available through live maps (e.g. Google Maps), other visualization techniques (e.g. GeoCommons (http://geocommons.com), Google Public Data Explorer (http://www.google.com/publicdata/directory) or applications that can be plugged into social networking sites.

There is a growing sentiment that governmental public health authorities should adopt and apply these Internet technologies to assess, protect and promote public health. While these innovations have a potential to meaningfully enhance some public health services (e.g. risk communication), its application to monitoring, detecting and investigating disease outbreaks for public health purposes, for example, is still limited. It is estimated that 90% of the data stored in the world today has been created in the past 2 years. In the 21st century, data are not just numbers; it is YouTube videos, Twitter posts (or 'tweets'), crowdsourcing information (which is engaging large groups of people to perform a task), etc.[1,2] One of the key advantages of online social media data, apart from the increasingly large data volumes, is that they are highly contextual and networked. For example, there is a strong spatio-temporal sentiment towards a new vaccine, whether positive or negative in nature. Risk factors—such as drug abuse, smoking, poor diet and exercise—and the associated diseases are often found to be clustered in the population.[3–6] A better understanding of social media and its health data province will help broaden the utility of social media in public health.

Social media is providing hope to answer some fundamental questions in the public health arena, including the identification of non-cooperative disease carriers ('Typhoid Marys'), adaptive vaccination policies, augmenting public health surveillance for early disease detection and creating disease situation awareness picture, updating or enhancing our understanding of the emergence of global epidemics from day-to-day interpersonal interactions, while engaging the public and communicating key public health messages. This article reviews the utility of social media for public health, with an emphasis on its application to disease surveillance. State-of-the-art innovations are discussed to illustrate capabilities these technologies enable, as well as knowledge gaps that limit the application of social media to public health work.

Public health, health data and the internet

Public health is 'the science and art of preventing disease, prolonging life and promoting health through the organized efforts and informed

choices of society, organizations, public and private, communities and individuals' (1920, C.E.A. Winslow). Information is core to public health, and health data are a corner stone. The timeliness of health data limits the availability of actionable public health information as the traditional route for the data moves from patient self-report to a physician, through diagnostic confirmations, and then from a physician or laboratory facility to a public health authority. Health data found on social media circumvent the traditional route by removing the 'middle man'. However, the middle man plays a critical role. They validate and can tell authorities something about the population they serve where generalizations of this information can be made with a level of confidence. That confidence is crucial in triggering a governmental action or intervention.

Public health officials, when looking back on the 2009 H1N1 event and Haiti Cholera outbreak, realized that social media could potentially indicate disease outbreak trends quicker and with higher reliability compared with traditional methods of public health reporting. Researchers and practitioners have begun to mine online data, Internet search behavior data and social network data in order to predict a variety of social, economic, behavioral and health-related phenomena. While much of the work has focused on predicting aggregate properties, such as the prevalence of seasonal influenza in a given location; in a country or a city (e.g. Google Flu Trends (http://www.google.org/flutrends), Monitoring Dengue activity using Internet search (http://www.google.org/denguetrends), SickCity, Sick weather ('How Are You Feeling Today?' http://www.sickweather.com), MappyHealth (http://mappyhealth.com, etc.), other work has focused on the health of an individual (e.g. Quantified Self, http://quantifiedself.com) and evaluating noisy and incomplete social data.[7–9]

The potential and challenges of social media in disease surveillance and public health interventions

Realizing the potential of social media in public health can provide timely and relevant information of public health importance. For example, social media has shown that it can be useful when seeking timely and reliable data on the spread or severity of influenza, west nile virus and meningitis.[10–12] Analyzing and disseminating timely information can improve public access to public health surveillance information. As we utilize open-source information, the operating cost can be extremely low while providing rich online features such as data-mining, categorizing, filtering, visualizing, timely sharing and predicting data on epidemics.[13] The transparency of social media may help demonstrate the value of openness in disease reporting, which may have 'spillover effects' on traditional public

health surveillance systems. However, social media does not represent a revolution in how we conduct disease surveillance. Current high-quality public health surveillance already utilizes multiple sources of information to gain a better understanding of the incidence and distribution of disease. For example, influenza surveillance may include laboratory-based virological surveillance, sentinel syndromic surveillance (e.g. tracking emergency room visits, school-based absenteeism reports, etc.) and evaluation of mortality trends for pneumonia and influenza, which taken together may provide a more complete picture of disease risk or impact.

Social media can be used effectively for engaging the public and communicating key public health messages. According to a survey conducted on 4033 clinicians by QuantiaMD and the Care Continuum Alliance in September 2011, nearly 90% of physicians reported using at least one social media site personally. One such example is Dr Majd Isreb, MD, FACP, FASN; a doctor specializing in Nephrology, Hypertension, Dialysis and Renal Transplant, who setup a professional Facebook page (https://www.facebook.com/pages/Majd-Isreb-MD-FACP-FASN/328625 408598). Dr Isreb posts daily to this page the latest scientific facts, encourages physical fitness, activity and spreads public health messages to patients and community members. Recent examples of his postings include messages such as: 'Please try to avoid drinking sport drinks (like Gatorade, etc.) for hydration, unless you are participating in a marathon, iron man or you are a professional athlete. They have lots of sugar and will lead to obesity for the 'sedentary' person!', and 'If you are cooking, garlic salt has less sodium than regular salt'.

During the 2009 H1N1 event, the US Centers for Disease Control and Prevention (CDC) turned to its Facebook page to educate the public about the disease and the importance of vaccination. This provided CDC the opportunity to engage with the public through a dialogue, spread public health messages and quickly correct any misinformation. Spearheaded by CDC Director, Dr Tom Frieden, engagement through social media became part of his communication strategy with the public. He is active on Twitter (@DrFriedenCDC), not only to communicate public health messages but also to engage citizens via hosting live Twitter chats. For example, when CDC responded to queries from Facebook and Twitter on health concerns regarding contracting H1N1 virus from eating pork, CDC was able to immediately answer questions and clear any confusion, rumors or misinformation.

Social media also has tremendous potential in assessing health behaviors and health sentiments or rumors. In a recent study, Salathé and Khandelwal[3] assessed the spread of vaccination sentiments from person to person during the unfolding of the H1N1 pandemic. They found that anti-vaccination sentiments could reliably be assessed across time and space, and that those sentiments seemed to cluster in certain parts of the

online networks. Further analysis indicated that negative sentiments spread more effectively than positive ones. As behavioral interventions are becoming increasingly important in public health, the potential of social media to study person-to-person transmission of health behaviors and sentiments in very large populations is unparalleled, and offers a clear benefit that traditional sources rarely provide.

Perhaps the most significant concern is the question of how to verify social media information, as the validation of large, noisy data sets poses enormous challenges. Public health officials typically have reservations about integrating social media as it could be an additional burden to their surveillance responsibilities. However, one preliminary way of analyzing social media data is through cross validation with other sources, which may help distill rumors early. A related concern with this type of approach is the risk of spreading rumors across multiple social networks or a malicious actor gaming the system with false information, where validation becomes difficult. To address these concerns many systems require messages to be reviewed by a moderator (either before or after public dissemination), label reports clearly as community contributions and enable users to provide feedback and even corroboration of submissions, as has proven successful with Wikipedia. In addition, social media is, by nature, a venue for two-way information exchange where crowds can verify and evaluate the quality of information shared by other users.

There are certain pitfalls to mining social media. First, textual data can be difficult to classify and interpret since harvested data (e.g. a tweet) may not provide enough information and meaning to facilitate automatic classification. Also, while coding for geographic origins may resolve certain limitations, not all profile accounts on social networking sites contain geographic information and visible geographic information cannot be easily verified. It is thus worthwhile to explore data mining sources that track IP addresses or techniques to monitor social media activity on mobile phones. Most mobile phones are equipped with Global Positioning Systems (GPS) monitoring chips or can be easily attached to independent devices. This feature can provide the supplemental location information needed, but the challenge remains in that users may not provide explicit consent to share this information publicly. As the social media landscape continues to evolve rapidly, special attention must be given to potential demographic biases among users of any given service.

Furthermore, the problem of over-fitting is well known in the domain of machine learning, a key methodology used in mining large data sets such as those extracted from social media. It is easy to construct models that perfectly fit millions of data points from social media sources to aggregate statistics from public health sources (e.g. disease incidence curves), but these models then need to be tested for their predictability. Another approach to address the problem is to avoid the use of data from

official sources during the process of developing the models. For example, Salathé and Khandelwal[3] extracted vaccination sentiments by classifying a fraction of the data manually, and then they used machine learning algorithms trained on human-labeled data to evaluate the rest of the data set. Only after this process was the fully labeled data set correlated against CDC estimates of H1N1 vaccination rates per geographic region, identifying a strong correlation between sentiments and vaccination rates (in the expected direction, e.g. vaccination coverage were higher in regions with more positive sentiment).

In the early stages of introducing a social media platform, strong biases are to be expected, but the demographic composition of the user base can change at dramatic speeds. For example, while virtually unknown in the general population 5 years ago,[14–23] latest Pew Internet study (Twitter Use 2012, Aaron Smith and Joanna Brenner http://www.pewinternet.org/Reports/2012/Twitter-Use-2012.aspx) found that Twitter is now used by hundreds of millions of people, and the demographic composition of its user base in the USA has only a few remaining biases.

It should be noted, however, that limitations to social media are not absent from traditional surveillance systems either. For example, an average population density can be significantly different from surveillance coverage. Estimates of incidence can change markedly with changing case definitions, incidence of laboratory-confirmed disease can change with augmentation or restriction of clinical testing, changes in diagnostic test methodologies and syndromic surveillance systems can be subject to poor specificity and frequent false alarms.

Social media platforms and their utility for disease surveillance and public health interventions

Social media platforms are designed to track and document events or relief efforts. For example: medical and pharmaceutical shortages in Kenya, Uganda, Malawi and Zambia; humanitarian crises (the Syria Tracker Crisis Map https://syriatracker.crowdmap.com); oil spills, public diplomacy;[24] election intimidation; corruption; tornados; power outages; civil wars; disease outbreaks; and distributing vaccines, food and water requests after an earthquake or hurricane (e.g. Hurricane Sandy 2012).

During the 2010 Haitian earthquake, Ushahidi, InSTEDD's Riff and HealthMap platforms allowed individuals affected by this crisis to post-information on lost individuals and track disease activity in the community. Rescue organizations were then able to use the information to reunite families and public health officials were able to track diseases as they propagated though the community using these platforms. This

enabled citizens of Haiti, anyone with a mobile phone, to get involved in responding to the crisis by subscribing to and committing their location-based information. Based on our experience and those in this field, we have selected the following examples to illustrate the various platforms and tools currently deployed in the field:

Ushahidi

Ushahidi (http://www.ushahidi.com) is a small, non-profit technology company based in Kenya that provides free and open source software as a platform through which one can leverage crowdsourced reports to generate live maps. At the bleeding edge of new practices deploying live maps in conflict and crisis situations, Ushahidi first launched its platform in 2008, which was used [in combination with Short Message Service (SMS)] to document rising human rights violations and post-election violence in Kenya. In 2010, Ushahidi released Crowdmap, a public implementation of the Ushahidi platform available for free via the Internet. Anyone can construct a basic Crowdmap within minutes, centered upon the location and extent of one's choice, and generated with relevant, user-defined categories that best fit the context. The web-based Crowdmap allows users to crowdsource information from multiple channels, such as email, Twitter, YouTube videos, online news, syndicated feeds (such as Geocoded Really Simple Syndication (or GeoRSS)), webform or mobile apps, among others.[25] Ushahidi allows users to review and assess the data submitted via curation, in addition to, automated algorithms (Swift River) for scoring and filtering information based on the credibility of sources. Collecting contact information from the person reporting, when available, can enable system owners to contact the submitter to request additional details if a report raises particular interest. With an effective review and filtering process, Ushahidi can help avoid information overload and reduce rumors.

InSTEDD's riff

InSTEDD's Riff (http://instedd.org/technologies/riff) is an open source social networking platform designed to streamline the collaboration between domain experts and machine learning algorithms for detection, prediction and response to health-related events (such as disease outbreaks) or disasters (InSTEDD was fielded by Google in 2006). Riff consists of five processes: (i) data gathering, (ii) automatic feature extraction, data classification and tagging, (iii) human input, hypotheses generation and review, (iv) predictions and alerts output and (v) field confirmation

and feedback (see Fig. 1a and b). Riff synthesizes health-related event indicators from a wide variety of information sources [structured and unstructured (e.g. news, social media, blogs)] into a consolidated picture for analysis and maintenance of 'community-wide coherence'. This helps detect anomalies, visualize clusters of potential events, predict the rate and spread of a disease outbreak and provide decision-makers with tools, methodologies and processes to investigate the event.

On 17 January 2010, the Thomson Reuters Foundation used Riff to launch a first-of-its kind, free disaster-information service for the people of Port Au Prince, Haiti. The use of Riff enabled survivors of Haiti's
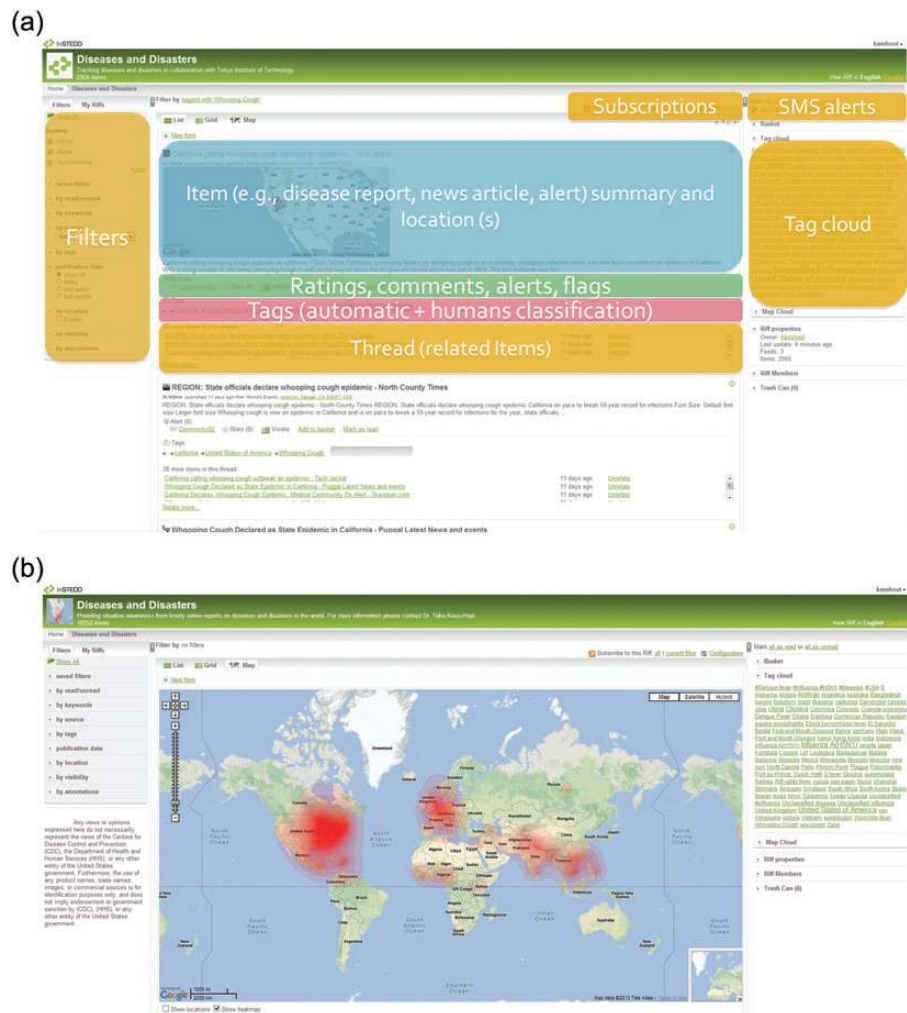
Fig. 1 (a) InSTEDD's Riff main features (e.g. disease reports, news articles, social media posts, locations, alerts, etc.). (b) InSTEDD's Riff: heat map of events detected from worldwide news media.

earthquake to receive critical information by text message directly to their phones, free of charge.[26,27]

HealthMap

HealthMap (http://healthmap.org) brings together disparate data sources to achieve a unified and comprehensive view of the current global state of infectious diseases and their effect on human and animal health. HealthMap relies on a variety of electronic media sources, including online news aggregators, eyewitness reports, expert-curated discussions and validated official reports. Through an automated process, updating 24/7/365, the system monitors, organizes, integrates, filters, visualizes and disseminates online information about emerging diseases in seven languages, facilitating early detection of global public health threats. HealthMap has demonstrated how informal sources of information, such as news media reports or Twitter postings, can be used complementarily with official data in an outbreak setting to get timelier (∼2 weeks earlier) estimates of disease dynamics when compared with government cholera cases reported in the first 100 days of the 2010 Haitian cholera outbreak. Estimates of the reproductive number ranged from 1.54 to 6.89 (informal sources) and 1.27–3.72 (official sources) during the initial outbreak growth period, and 1.04–1.51 (informal) and 1.06–1.73 (official) when Hurricane Tomas afflicted Haiti. HealthMap was created in 2006, and it is an established global leader in utilizing online informal sources for disease outbreak monitoring, with over a million users a year. HealthMap delivers timely intelligence on a broad range of emerging infectious diseases for a diverse audience including local health departments, governments and international travelers.[13,28–30]

FluNearYou

FluNearYou (FNY) (http://flunearyou.org) is a project developed by HealthMap, the American Public Health Association and Skoll Global Threats Fund, which allows the public to complete brief weekly surveys to help us all learn about the flu. The goal of the project is to show how timely flu data and cooperation across sectors can better prepare us for future pandemics. The site, administered by HealthMap, is free and accessible to everyone; however, currently those who wish to participate in the surveys must be US or Canada residents over 13 years of age (see Fig. 2). The FluNearYou website has integrated user-contributed data (via web or mobile phone entries) from over 43 000 volunteers, CDC Flu Activity data and Google Flu Trends ILI (Influenza-Like Illness) data.

Fig. 2 Flu near you (FNY).

Users are able to see the recent prevalence of symptoms near them, an animation of the increase in flu activity and data points since the beginning of the project (October 2011), links to local public health entities, and vaccine availability near them.[31] Google's Flu Vaccine Finder supplied the vaccine availability data originally, but this service is now run by HealthMap (http://flushot.healthmap.org) starting with the 2012–13 flu season.

MappyHealth

MappyHealth (http://www.mappyhealth.com) is a web-based application that uses open source components and provides a variety of analytical views of disease-oriented tweets providing high precision and good recall on the basis of someone's social ties and co-locations with other people, as revealed by their tweets. The first and most obvious challenge of using social media is that it is unstructured, in abundance, and very noisy. In June 2012 it was estimated that 400 million tweets per day are created with anticipation of almost 2 billion users on Twitter [The statistics on tweets per hour are relevant in light of studies demonstrating that between one and four tweets per hour is optimum for achieving the maximum click through (visibility) http://www.mediabistro.com/alltwitter/science-social-timing_b10473].[32,33]

MappyHealth had to address the following challenges as the platform was constructed: (i) what problem is being solved, (ii) how will users of the
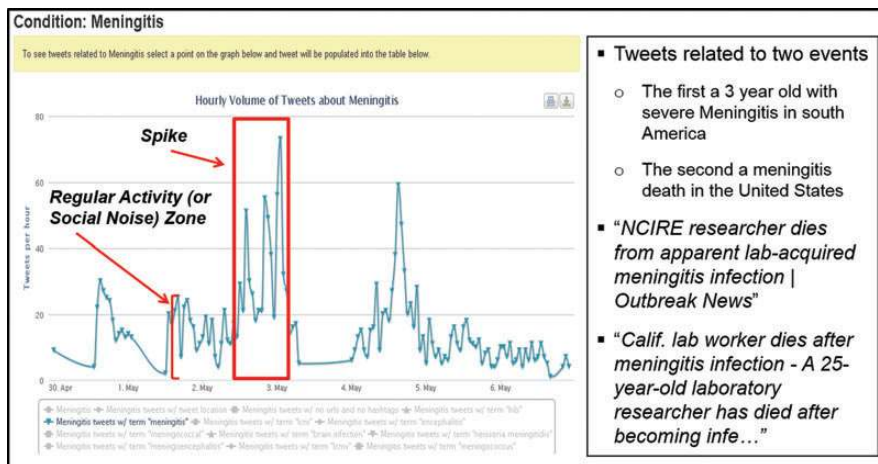
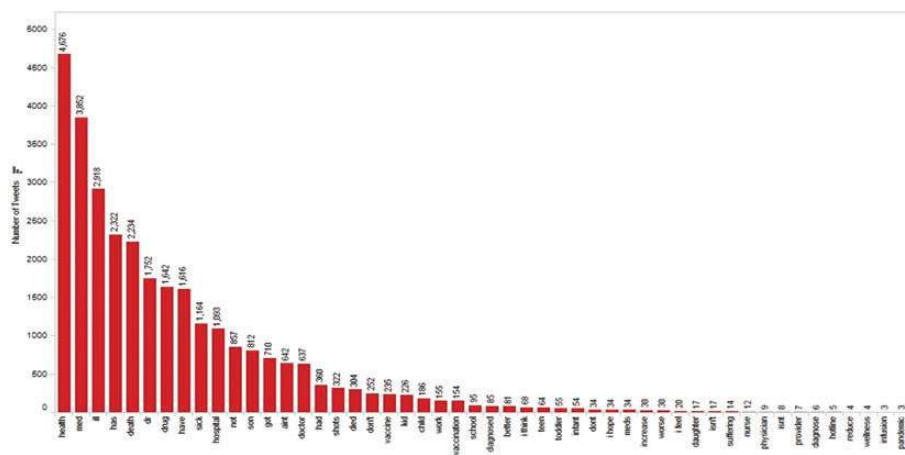**Fig. 3** Meningitis tweet activity, 30 April 2012 to 8 May 2012.

Fig. 4 Meningitis tweet qualifier activity, 19 September 2012 to 19 October 2012.

platform utilize the information provided, (iii) how best to visualize the data, (iv) what semantics will be meaningful, (v) what analytical intervals are pertinent and (vi) can the information be geo-located data. To address each of these challenges, MappyHealth started utilizing the Twitter streaming Application Programming Interface (API) to collect tweets associated with over 200 health terms placed into 26 categories. MappyHealth ingests only a fraction of the daily twitter volume at a rate of ~1–1.5 million tweets per day and in timely fashion. Currently, over 100 million tweets are publicly available on MappyHealth for public consumption or research. As Tweets are brought in the platform processes them to begin understanding the semantic meaning through tagging tweets with over 200 qualifiers such as sick, child, I have, etc. MappyHealth found that social analytics in the public health space can enable the ability to track disease, and in addition

to tracking disease can provide insights into social chatter allowing public health officials (and other health-care organizations) to monitor rumors, messaging and other forms of communications. This type of insight could be uniquely powerful as health care and public health organizations converge to engage patients in meaningful ways. For example, in 2012 the USA experienced two separate outbreaks of meningitis. The number of tweets associated with meningitis and what people are tweeting about correlate with the events occurring (see Figs 3 and 4). Certainly more analytics and validation is warranted to enhance future analytics and insight, but nonetheless it is hard to ignore this observed phenomenon.

The climate change collaboratory (Triple-C)

Triple-C is an interdisciplinary initiative to encourage and study discourse and critical debate that lead to a shared understanding of climate change issues on all political levels, ranging from inter-individual communication and local communities to global campaigns and treaties. By investigating communicative strategies and processes that function between disciplines and stakeholders, the Triple-C project aims to unearth hidden assumptions and misconceptions about climate change, contribute to a mutual understanding of existing problems and suggest priorities for research and policy development. Participants of the collaboratory benefit from a synergy of skills and resources, the constitution and dynamic maintenance of shared knowledge, flexible and non-hierarchical modes of cooperation and mechanisms for distributed decision-making.

Environmental web resources such as documents and best-practice examples are often being created through processes of cooperation and social exchange. They depend on and benefit from a synergy of skills, the dynamic maintenance of shared knowledge, flexible and non-hierarchical portfolios of services and distributed decision-making. Triple-C is a project funded by the Austrian Climate and Energy Fund within the Austrian Climate Research Program. It aims to strengthen the relations between Austrian scientists, policy-makers, educators, environmental NGOs, news media and corporations—stakeholders who recognize the need for adaptation and mitigation, but differ in world views, goals and agendas. The Collaboratory manages expert knowledge and provides a social networking platform for effective communication and collaboration. It assists networking with leading international organizations, bridges the science-policy gap and promotes rich, self-sustaining community interaction to translate knowledge into coordinated action. Innovative survey instruments in the tradition of games with a purpose create shared meaning and leverage networking platforms to capture indicators of environmental attitudes, lifestyles and behaviors.[34,35]
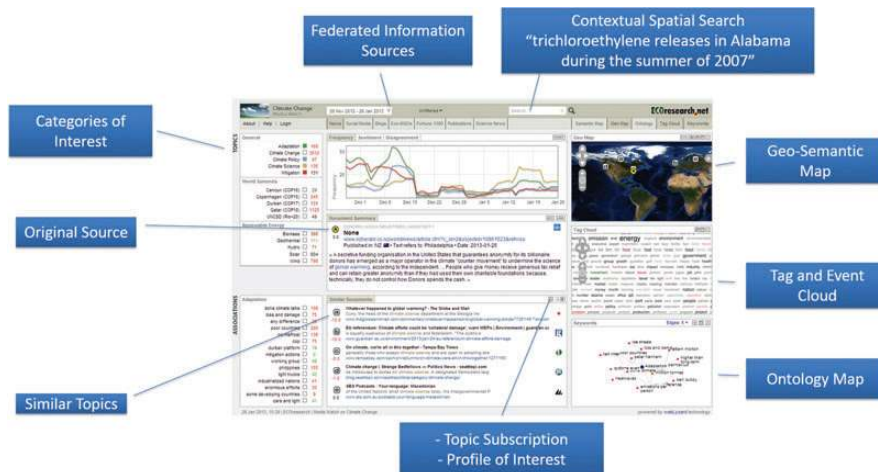
Fig. 5 Media watch on climate change collaboration (Triple-C) platform.

Triple-C recognizes and supports the social construction of meaning via distributed information services that aim to improve the quality of decisions, build trust and help resolve conflicts among competing interest. It provides matchmaking services for *ad hoc* team composition and a range of web-enabled communication and collaboration tools. Facilitating the collaboration between stakeholders requires a tight integration of heterogeneous services. Collaborative ontology building ensures that recent findings are understood by all members of a virtual community. Triple-C draws upon the lessons learnt from building the Media Watch on Climate Change,[36] which is available online at http://www.ecoresearch.net/climate (see Fig. 5). This social network and news platform provides geographic and semantic visualizations based on multiple coordinated view interface. It provides communication and collaboration tools such as messaging services, Wikis, web-based discussion forums, multi-language support and a layered security model to distinguish between public and private information. Geographic mapping plays a central role, using the virtual globe technology of NASA World Wind to integrate different types of data objects (documents, best-practice examples, expert profiles, social media, news, blogs, etc.), and put them into a regional context.

Crisis Tracker

Crisis Tracker (http://ufn.virtues.fi/crisistracker) is the first platform to use timely social media as a structured information source during mass disasters (see Fig. 6). Crisis Tracker collects social media, such as Twitter, data on regular basis and then applies text mining (agnostic to spoken languages) and event detection algorithms adaptable to crowd input (or

Crisis Tracker Underlying Process: Artificial Intelligence for Disaster Response (AIDR)
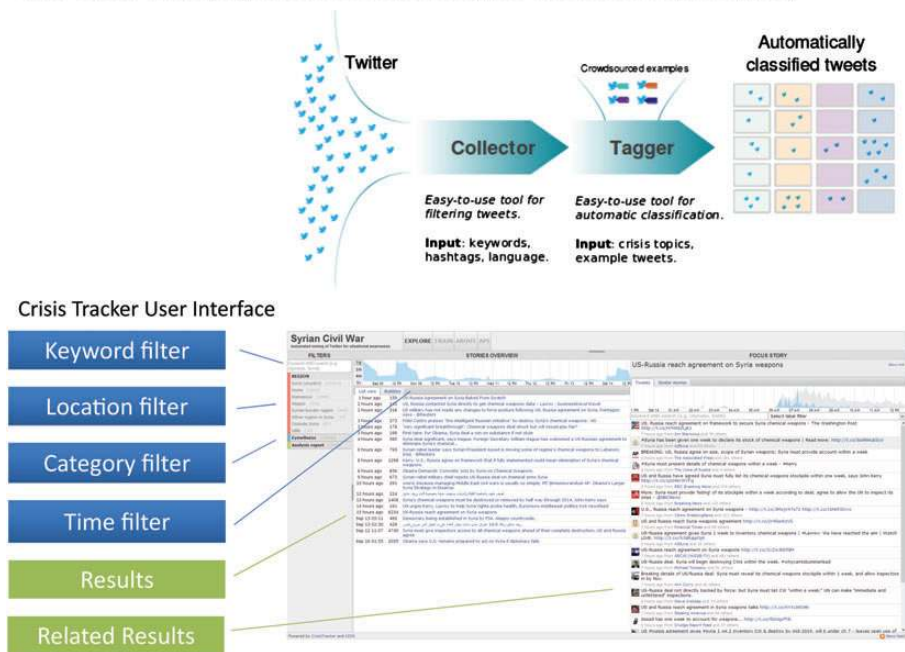

Crisis Tracker User Interface

Fig. 6 Crisis tracker user interface.

curation). In addition to displaying raw tweets, the system provides an intermediary level of summaries that retain details within reports. Crisis Tracker is scalable for use during mass disasters and conflicts for tracking events as they unfold, which makes it unique compared with other similar platforms such as: (i) Sahana (http://sahanafoundation.org) and VirtualAgility OPS Center (VOC) (http://www.virtualagility.com). These two systems often integrate raw social media feeds, but lack capabilities for distilling and handling situations when activity is exceptionally high; (ii) Ushahidi which its effectiveness depends entirely on the size, coordination and motivation of crowds which adapts well to needs of specific disasters, but is difficult to scale to match information inflow rates during very large events; (iii) Twitcident (http://twitcident.com), which works only with geo-tagged tweets (~1% of all posted messages), employs classification algorithms (spoken language specific and requires training for every time a new concept is introduced) to extract situation awareness information during small-scale crisis response, such as music festivals or factory fires. This system, however, is not built to monitor large and complex events with multiple parallel storylines or for emerging events or threats, such as a novel disease outbreak and (iv) EMM NewsBrief (http://emm.newsbrief.eu) mines and clusters mainstream news media from predetermined sources in a wide range of spoken languages, with new
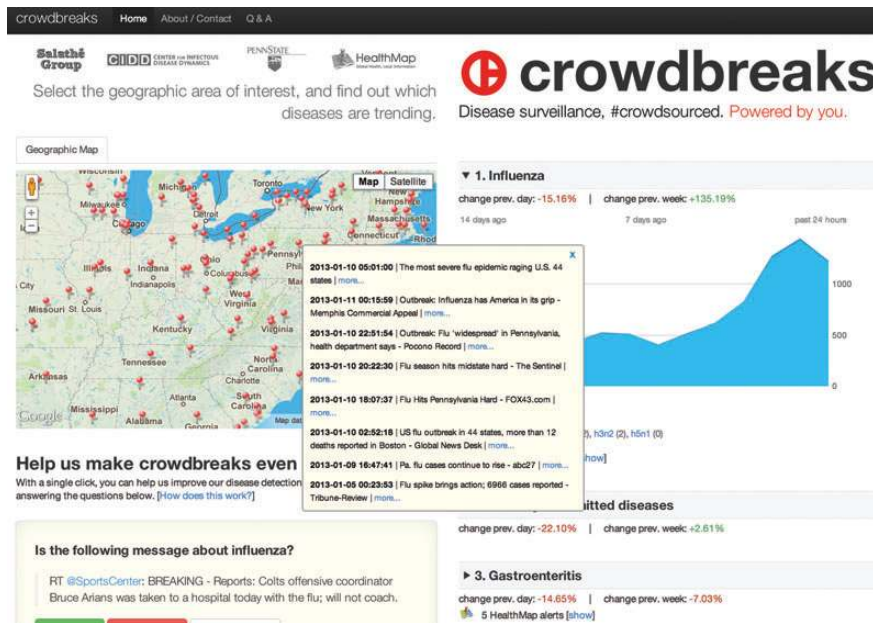
Fig. 7 Crowdbreaks user interface.

summaries updated every 10 min, but has not been extended to handle social media.

Crowdbreaks

Crowdbreaks (http://crowdbreaks.com) is a crowdsourced disease surveillance system that collects tweets containing disease-related keywords (see Fig. 7). It determines the location of origin of tweets and employs machine learning algorithms to assess the relevance of each tweet. Crowdbreaks uses crowdsourcing to label the tweets by inviting site visitors to answer a simple question about a random tweet from the raw data set. These answers are then used as training data for the machine learning algorithms. It integrates HealthMap data, which provides event-based epidemic intelligence in addition to the Twitter data. While still in beta, Crowdbreaks offers a glimpse of a future where noisy, unstructured social media data are not only provided by the crowd, but also assessed and curated by the crowd for their relevance to the issue at hand.

The role of mobile phones

The near-ubiquity of mobile phones worldwide and enhanced access to the Internet over the past few years has allowed collaboration among broader

cross section of the health care and public health community and the public. With 70% of the world's population carrying mobile phones, individuals increasingly have the technological means to document and publicize their health status. This rise in adoption of mobile phones and the Internet, in both industrialized and developing countries, has provided additional opportunities in crowdsourcing. Mobile phones hold particular promise for this type of opportunity because they can be used as point-of-care devices, function in remote locations and are readily carried and used at any time.[27] Additionally, because smartphone applications include capability to register GPS coordinates, as mentioned above, verification of the proximity of the reporter to the location in question can also be used as a validation and verification tool.

Selected mobile applications or platforms for health

OpenData kit (http://opendatakit.org), a suite of open-source tools that make use of existing mobile, allows Kenyan medical workers to track and upload patient medical information directly into the medical record system using their mobile phones. InSTEDD's GeoChat (http://instedd.org/technologies/geochat) is another open source technology, which allows team members in emergency situations to 'connect, visualize, report, receive and coordinate data and information' using mobile phones. HealthMap released a mobile application following the first wave of the H1N1 pandemic called Outbreaks Near Me (http://www.healthmap.org/outbreaksnearme), which asks its users to contribute reports of influenza-like illness using smart phone. Outbreaks Near Me received ~110 000 downloads and collected over 2400 submissions.[27] Propeller Health (previously known as Asthmapolis) (http://propellerhealth.com) maps asthma triggers and identifies the severity of asthma attacks when patients use inhalers equipped with special trackers. Propeller Health is meant to track and further the medical knowledge on environmental asthma triggers.

After the 2010 earthquake and cholera outbreak in Haiti, a research group at the Karolinska Institute in Stockholm tracked population movements using anonymous SIM (subscriber identification module) card data from mobile phone providers. This led to an effort to mine data from phones during disasters to track population movements for relief agencies and to detect disease outbreaks.[37] Mobile phones can be a great tool to enable crowdsourcing for public health. More than 90% of Kenyans use mobile phones, giving scientists a powerful tool to track how malaria spread, discover regional routes around Lake Victoria (which serve as the major disease corridors for the parasite) and towns along the routes that are hot spots for transmitting malaria into megacities, such as Nairobi.[38]

Conclusion

The general consensus in the health care and public health community is that we need to pay attention to social media.[39] Social media can be a great tool to help us study person-to-person spread of communicable diseases and behaviors, in addition to better understanding of non-communicable diseases. These diseases include, depression, type-II diabetes, cardiovascular or pulmonary illnesses, which pose a substantial public health risk and are typically associated with certain behavioral factors. Also, social medial can help shorten the length of time it takes to detect disease outbreaks and improve responses, allowing health care and public health agencies to engage and communicate with the public. Another unexpected byproduct of social media is that it made it easier for health-care agencies and governments to share data with their citizens and better understand how sentiments or rumors spread in order to engage with the appropriate public health messages and to deploy the right interventions. Overall, however, social media serves primarily as a complementary tool, rather than a replacement for either traditional population monitoring efforts or existing new-generation Internet systems. Additionally, as the usage of social media data to monitor health and disease dynamics grows and matures over time, validation will always be a key issue that needs to be given careful attention. More work is warranted to validate and prove the value of social media as a tracking or prediction tool. The background noise will increase exponentially when information becomes more available, and with it, rumors and half-truths.

References

1 Nicholas P. *The Missing Step: Statistical Inference from Big Data*. Mathematical Biosciences Institute, 2012. http://mbi.osu.edu/2012/10thmaterials/jewell.pdf (21 September 2013, date last accessed).

2 Jewell NP. Counting civilian casualties. In: Herzberg AM (ed.). *Statistics, Science and Public Policy XVII, Democracy, Danger and Dilemmas*. Kingston: Queen's University, 2013.

3 Salathé M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLoS Comput Biol* 2011;**7**:e1002199.

4 Salathé M, Bengtsson L, Bodnar TJ *et al*. Digital epidemiology. *PLoS Comput Biol* 2012;**8**:e1002616.

5 Mokdad AH, Marks JS, Stroup DF *et al*. Actual causes of death in the United States, 2000. *JAMA* 2004;**291**:1238–45.

6 Schuit AJ, van Loon AJM, Tijhuis M *et al*. Clustering of lifestyle risk factors in a general adult population. *Prev Med* 2002;**35**:219–24.

7 Google flu trends: 'how does this work?' http://www.google.org/flutrends/about/how.html (21 September 2013, date last accessed).

8 Madoff LC, Fisman DN, Kass-Hout T. A new approach to monitoring dengue activity. *PLoS Negl Trop Dis* 2011;**5**:e1215. http://www.plosntds.org/article/info%3Adoi%2F10.1371%2Fjournal.pntd.0001215 (21 September 2013, date last accessed).

9 Sadilek A, Kautz HA, Silenzio V. Predicting disease transmission from geo-tagged micro-blog data. In: AAAI. 2012. http://www.aaai.org/ocs/index.php/AAAI/AAAI12/paper/viewFile/4844/5130 (21 September 2013, date last accessed).

10 Wilson K, Brownstein JS. Early detection of disease outbreaks using the internet *CMAJ* 2009;**180**:829–31.

11 MappyHealth, Meningitis tweet data visualization (19th September–6 October 2012). http://socialhealthinsights.com/2012/10/meningitis-tweet-data-visualization-sept-19th-oct-6th-2012 (21 September 2013, date last accessed).

12 MappyHealth, West Nile Virus; visual analysis of tweet activity (1 August 2012 to 8 September 2012). http://socialhealthinsights.com/2012/09/west-nile-virus-analytical-analysis-of-tweet-activity-aug-1st-2012-to-sept-8th-2012 (21 September 2013, date last accessed).

13 Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection—harnessing the web for public health surveillance. *N Engl J Med* 2009;**360**:2153–2155, 2157.

14 Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. *PLoS One* 2011;**6**:e19467.

15 Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of tweets during the 2009 H1N1 outbreak. *PLoS One* 2010;**5**:e14118.

16 Moher D, Liberati A, Tetzlaff J *et al*., PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;**151**:264–9.

17 Achrekar H, Avinash G, Lazarus R *et al*. Predicting flu trends using Twitter data. International Workshop on Cyber-Physical Networking Systems (CPNS) in conjunction with INFOCOM 2011 CPNS Shanghai, China, 2011, http://www.cs.uml.edu/~hachreka/SNEFT/images/CPNS2011_pdf.pdf (21 September 2013, date last accessed).

18 Chen L, Achrekar H, Liu B *et al*. Vision: towards real time epidemic vigilance through online social networks: introducing SNEFT. ACM Workshop on Mobile Cloud Computing & Services: Social Networks and Beyond (MCS), San Francisco, USA, 2010. ACM 978-1-4503-0155-8. http://www.cs.uml.edu/~bliu/pub/sneftFinal.pdf (21 September 2013, date last accessed).

19 Corley CD, Mikler AR, Singh KP *et al*. Monitoring influenza trends through mining social media. In: International Conference on Bioinformatics and Computational Biology. Las Vegas, NV, 2009.

20 Corley CD, Cook DJ, Mikler AR *et al*. Text and structural data mining of Influenza mentions in web and social media. *Int J Environ Res Public Health* 2010;**7**:596–615.

21 Corley CD, Cook DJ, Mikler AR *et al*. Using web and social media for Influenza surveillance. *Adv Exp Med Biol* 2010;**680**:559–64.

22 de Quincey E, Kostkova P. Early warning and outbreak detection using social networking websites: the potential of twitter. In: *Electronic Healthcare*, pp. 21–4. Berlin Heidelberg: Springer, 2010.

23 Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. *Am J Prev Med* 2011;**40**:S154–8.

24 Hend Alhinnawi, e-Democracy: Egypt's 18 Day Revolution. http://www.slideshare.net/kasshout/edemocracy-egypts-18-day-revolution (21 September 2013, date last accessed).

25 Meier P. 2012. 'Ushahidi: Blog'. Crisis mapping Syria: automated data mining and crowdsourced human intelligence. Ushahidi, 27 March. http://blog.ushahidi.com/index.php/2012/03/27/crisis-mapping-syria (4 November 2012, date last accessed).

26 Kass-Hout TA, di Tada N. International system for total early disease detection (InSTEDD) platform. *Adv Dis Surveill* 2008;**5**:108.

27 Freifeld CC, Chunara R, Mekaru SR *et al*. Participatory epidemiology: use of mobile phones for community-based health reporting. Health in action. *PLoS Med* 2010;**7**:e1000376. http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1000376.

28 Brownstein JS, Freifeld CC, Reis BY *et al*. Surveillance Sans Frontières: Internet-based emerging infectious disease intelligence and the HealthMap project. *PLoS Med* 2008;**5**:e151.

29 Freifeld CC, Mandl KD, Reis BY *et al*. HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. *J Am Med Inform Assoc* 2008;**15**:150–7.

30 Global Health: Special Focus on HaitiChunara R, Andrews JR, Brownstein JS. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian Cholera outbreak. *Am J Trop Med Hyg* 2012;**86**:39–45.

31 Anna Tomasulo; Meet Collaborate | Activate Finalist FluNearYou, 27 September 2012. http://collaborateactivate.com/blog/flunearyou (21 September 2013, date last accessed).

32 Bennett S. Twitter now seeing 400 million tweets per day, increased mobile Ad revenue, says CEO, 2012. http://www.mediabistro.com/alltwitter/twitter-400-million-tweets_b23744#more-23744 (21 September 2013, date last accessed).

33 Social Health Insights. Meningitis data visualization blog post, 2012. http://socialhealthinsights.com/2012/10/meningitis-tweet-data-visualization-sept-19th-oct-6th-2012 (21 September 2013, date last accessed).

34 Rafelsberger W, Scharl A. Games with a purpose for social networking platforms. In: 21st ACM Conference on Hypertext and Hypermedia, Torino, Italy: Association for Computing Machinery, 2009,193–7.

35 Abbasi DR. *Americans and Climate Change: Closing the Gap between Science and Action*. New Haven: Yale School of Forestry and Environmental Studies, 2006.

36 Hubmann-Haidvogel A, Scharl A, Weichselbraun A. Multiple coordinated views for searching and navigating web content repositories. *Inform Sci* 2009;**179**:1813–21.

37 Bengtsson L, Lu X, Thorson A *et al*. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. *PLoS Med* 2011;**8**:e1001083.

38 Wesolowski A, Eagle N, Tatem AJ *et al*. Quantifying the impact of human mobility on malaria. *Science* 2012;**338**:267–70.

39 Thackeray R, Neiger BL, Smith AK *et al*. Adoption and use of social media among public health departments. *BMC Public Health* 2012;**12**:242.