COLUMN / CHRONIQUE

Social media trends for health librarians: a primer on using social media for clinical disease surveillance

Dean Giustini

Introduction

This column introduces health librarians to the relatively new field of *clinical disease surveillance*. As such, it looks at how social media is currently being used around the world to track illnesses in real time by using key international websites such as Google Flu and HealthMaps. Clinical surveillance is a term that refers to an ongoing collection and analysis of health data about diseases that may pose a threat to public health [1]. Many international institutions, such as the World Health Organization and the US Centers for Disease Control [2], use databases and computer systems to track flu outbreaks, infection rates of the human immune deficiency virus (HIV), and bioterrorism threats. Further, according to the CDC

"...'syndromic surveillance' applies to health-related data that precedes diagnosis and signals a sufficient probability of an outbreak to warrant further public health response...though historically syndromic surveillance has been utilized to target the investigation of potential cases, its utility for detecting outbreaks associated with bioterrorism is increasingly being explored by public health officials."[1]

In 2012, clinical information systems that automate part of the early warning process for emerging illnesses intersect with the practice of informatics [3]. Early warning systems have found favour with institutions that oversee health services accreditation such as the US Joint Commission and those involved in medication error surveillance. In addition, surveillance of clinical information is a useful aid in adverse drug event reporting and in the broader field of pharmacoeconomics [3]. Any data, in fact, providing predictive or forecasting value to the clinical setting will be used by decision-makers and administrators in charge of health utilization. With respect to social media, some health informaticians regard its use as novel and even important to their work [4]; still, others are skeptical. Traditional techniques for detecting changes in health in human populations historically have focussed on data stored by clinical labs and epidemiological agencies, and perhaps even by public health agencies. In 2006, the Canadian health informatician Gunther Eysenbach at the University of Toronto discussed the notion of using search queries in Google as a kind of global sentinel for flu outbreaks [4]. In 2009, he expanded on the use of using search queries data in Google for monitoring public health (which he refers to as 'infodemiology' as well as 'infoveillance'). [5] The important message for health librarians is that search engines as well as social media tools such as Facebook and Twitter are increasingly used in data mining and as rich sources of information to follow emerging threats to the public health.

Tracking flu in real time

Researchers at Harvard have examined some of the pitfalls of using health data culled from social networking sites to ascertain flu outbreaks [6]. One of the challenges faced by epidemiologists and public health officials is coming up with reliable predictive models in established social networks for those considered to be "influential spreaders" of disease, and consequently whether generalizing these models to larger populations is scientifically sound [7-10]. In 2011, Eysenbach analyzed Twitter content from the H1N1 outbreak between 2008 and 2011 and assessed its utility in clinical surveillance [11]; his article generated a lot of criticism. In a different twist, Eke has used Twitter and tweets sent over the web in order to track dental pain [12], with mixed results. That said, a few websites that mine social media datasets to monitor influenza outbreaks are worth examination.

Flu Detector <http://geopatterns.enm.bris.ac.uk/epidemics/> was developed by the Intelligent Systems Laboratory at the University of Bristol in the United Kingdom. The site says that "...Flu Detector uses the content of Twitter to 'nowcast' flu rates in several UK regions" providing

Dean Giustini. UBC Biomedical Branch Library, Gordon and Leslie Diamond Health Care Centre, 2775 Laurel Street, Floor 2, Vancouver, BC V5Z 1M9, Canada.

Corresponding author (e-mail: dean.giustini@gmail.com).

time-sensitive information "in real time" about human influenza rates.

Google Flu Trends Canada http://www.google.org/flutrends/ca/#CA identifies flu-related searches on Google as a measure of flu activity in specific regions. In a study published in *Nature* [13], Google's flu findings were said to be more accurate than data gathered through traditional methodologies by the Centers for Disease Control and Prevention (CDC) by as much as 1 to 2 weeks. An in-depth analysis of influenza-related content on the blogosphere identified similar findings.

Influenzanet <http://www.influenzanet.eu/> is a surveillance system detecting so-called self-reported information of flu symptoms. In 2011, ten countries participated in a European survey via each country's Influenzanet partner site e.g., Gripenet (http://www.gripenet.pt/) in Portugal, Influweb (http://www.influweb.it/) in Italy, and Flusurvey http://flusurvey.org.uk/ in the United Kingdom. In Australia, a comparable Flutracking project encouraged participants to complete an online survey about their flu-like symptoms. By 2012, there are more than 10 000 weekly participants who have collectively completed 875 500 surveys, and have had 117 years off work due to influenza-like illness [14].

Pros and cons of surveillance tools

There are several advantages of using Internet-based surveillance methods to track emerging illnesses or pandemics. In addition to the rich source of real time data on the web, the retrieval and collection of this information is relatively easy and inexpensive. The challenges of using online surveillance data, however, can be equally potent: there is a real potential for social media data to produce skewed or biased datasets, and there are ethical concerns associated with using this health information. For example, participants cannot opt out of the data gathering process as this information is mined from public sites and search engines – and generally speaking, the public is unaware that health information is used in this way without their explicit permission. Researchers have debated whether traditional statistical methods should be used to analyze social data from Twitter and Facebook, or whether new statistical approaches will be needed. Even though social network analysis is an established discipline, there are numerous challenges in aggregating millions of "unstructured tweets" and blogposts from social media websites, especially with respect to the reliability and generalizability of this information. More research is clearly needed.

Surveillance mashups

According to Cho [15], mashups can be defined as hybrid tools that take the functionality of two or more sources of information in health, merging aspects of each, in order to create a third source. The ability to create mashups on the web has provided some interesting opportunities for public health organizations involved in clinical surveillance. For example, mapping data can be combined using Google Maps API and epidemiological information can be mined from the Internet through the use of various social media sources. One example, HealthMap <http://www.healthmap.org/en/>, tracks emerging diseases in real time using multiple data sources such as the World Health Organization and Euro Surveillance. HealthMap also provides pinpoint visuals on online maps to see where diseases are most prevalent, information that is shared on Healthmap's Twitter feed. Other surveillance websites include the Avian Flu Mashup <http://www. nature.com/avianflu/google-earth/index.html>, which was developed in 2006 by the folks at Nature publishing. This award-winning site tracks the spread of H5N1 using the Google Maps API. The site aims to control the spread of Avian flu by drawing on the many lessons from the 2009 pandemic, and by following current news of avian flu outbreaks throughout the world. Finally, WhoisSick? <http://whoissick.org/sickness/> is one of the older health mashups as it was created in early 2006. Its mission is to provide current and local sickness information in the Los Angeles area. With the belief in the power of people and a faith in user-generated data, this epidemiological mashup links "sickness data" with Google Maps. WhoIsSick? is intended for health consumers who want to see disease outbreaks online as they occur.

Surveillance information and websites

US Centers for Disease Control (CDC)

The CDC has one of the most impressive social media campaigns of any public health agency in the world. In the last few years, it has introduced Facebook, YouTube, Twitter and other social media tools to disseminate public health information, and publishes a toolkit for those who want to use best practices within their organization. Using social media tools has become an effective way to expand the reach and engagement of health consumers, and to broadcast credible health information. A number of surveillance documents from the CDC are also listed below and serve as a contrast to using the social media tools mentioned to stay current with what is happening at the CDC:

- CDC's Social Media Campaign
- CDC's Behavioral and Clinical Surveillance Branch
- Morbidity and Mortality Weekly Report
- Emergency Preparedness and Response Surveillance
- Seasonal Influenza & Emerging Infections Network

Public Health Agency of Canada (PHAC)

The PHAC disseminates health information via e-mail, podcasts, and RSS feeds, to name a few social tools. Its newest feature is a widget that provides H1N1 updates and general PHAC news. A widget is an application that displays content from a website directly on your home page. By embedding the widget code on your site, blog or social media profile, live content from PHAC will automatically be displayed. A number of surveillance documents from PHAC are also listed below, and serve as a contrast to using the social media tools mentioned.:

- Public Health Agency of Canada (PHAC) Mobile and Social Media Tools
- Canada Communicable Disease Report weekly

- Canadian Integrated Public Health Surveillance
- Canadian Perinatal Surveillance System
- Chronic Disease
- Disease Centre and Control Guidelines
- Disease Surveillance On-line
- FluWatch
- Surveillance
- Travel Health Notices
- West Nile Virus Monitor

Conclusion

With the proliferation of social media tools and activities on the web, public health officials, epidemiologists and health librarians are starting to use these tools in their data mining and literature surveillance activities. As more research is published, and public health organizations seek to mine this information from social websites, health librarians will want to examine the possibility of providing expertise on social networks as well as in data curation and mining services.

References

- 1. National Research Council. Global Infectious Disease Surveillance and Detection: Assessing the Challenges – Finding Solutions, Workshop Summary. Washington, DC: National Academies Press, 2007.
- 2. Centers for Disease Control and Prevention. H1N1 Web and social media metrics; cumulative data report; 2009.
- 3. Hersh WA. Information retrieval: a health and biomedical perspective. *Springer*, 2009.
- 4. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. *AMIA Annu Symp Proc.* 2006:244–248.
- Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res. 2009;11(1):e11. doi:10.2196/jmir.1157.
- Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection-harnessing the web for public health surveillance. *N Engl J Med.* 2009;360(21):2153–2157. doi:10.1056/NEJM p0900702.
- 7. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clinical Infect Dis.* 2009;49:1557–1564. doi:10.1086/630200.
- 8. Bravata DM, McDonald KM, Smith WM, Rydzak C, Szeto H. Systematic review: surveillance systems for early detection of bioterrorism-related diseases. *Ann Intern Med.* 2004 Jun 1;140(11):910–922.
- 9. Corley CD, Cook DJ, Mikler AR, Singh KP. Using web and social media for influenza surveillance. *Adv Exp Med Biol.* 2010;680:559–564.
- Christakis NA, Fowler JH. Social network sensors for early detection of contagious outbreaks. *PLoS One.* 2010 Sep 15; 5(9):e12948. doi:10.1371/journal.pone.0012948.
- 11. Eysenbach G. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional

metrics of scientific impact. *J Med Internet Res.* 2011 Dec 19; 13(4):e123. doi:10.2196/jmir.2012.

- 12. Eke PI. Using social media for research and public health surveillance. *J Dent Res.* 2011;90(9):1045–1046. doi:10.1177/0022034511415277.
- Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature*. 2009;457(7232): 1012–1014. doi:10.1038/nature07634.
- 14. FluTracking http://www.flutracking.net/
- 15. Cho A. An introduction to mashups for health librarians. JCHLA/JABSC. 2007;28(1):19–21. doi:10.5596/c07-007.

Other notable clinical surveillance research

- Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS One.* 2010;5(11):e14118.
- Chunara R, Andrews JR, Brownstein JS. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. *Am J Trop Med Hyg.* 2012 Jan;86(1):39–45.
- Christley RM, Pinchbeck GL, Bowers RG, Clancy D, French NP. Infection in social networks: using network analysis to identify high-risk individuals. *Am J Epidemiol*. 2005;162:1024–31.
- Dreesman J, Denecke K. Challenges for signal generation from medical social media data. *Stud Health Technol Inform.* 2011; 169:639–41.
- Jenkins WD, Wold B. Use of the Internet for the surveillance and prevention of sexually transmitted diseases. *Microbes and Infection*. 2011;14(5):427–37.
- Khan S, Maclean CD, Littenberg B. The effect of the Vermont Diabetes Information System on inpatient and emergency room use: results from a randomized trial. *Health outcomes research in medicine*. 2010;1(1):e61–6.
- Kilbridge PM, Campbell UC. Automated surveillance for adverse drug events at a community hospital and an academic medical center. J Am Med Inform Assoc. 2006;13(4):372–7.
- Lampos V, De Bie T, Cristianini N. Flu detector tracking epidemics on Twitter. *Lecture Notes in Computer Science*. 2010;6323:599–602.
- Newkirk RW. The potential capability of social media as a component of food safety and food terrorism surveillance systems. *Foodborne Pathogens and Disease*. 2012;9 (2):120–4.
- Reynolds BJ. Building trust through social media. CDC's experience during the H1N1 influenza response. *Mark Health Serv.* 2010;30(2):18–21.
- Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: Twitter and antibiotics. *Am J Infect Control*. 2010;38(3):182–8.
- Tilston NL, Eames KT, Paolotti D, Ealden T, Edmunds WJ. Internet-based surveillance of Influenza-like-illness in the UK during the 2009 H1N1 influenza pandemic. *BMC Public Health*. 2010 Oct 27;10:650.