



Note: This provisional PDF corresponds to the article as it appeared upon acceptance, prior to any copyediting. The final, formatted and edited version will soon be made available in PDF and full text (HTML) formats.

The Landscape of International Event-based Biosurveillance

Emerging Health Threats Journal 2009, **3**:e3 doi: 10.3134/ehjtj.10.003

DM Hartley (hartley@isis.georgetown.edu)

NP Nelson

R Walters

R Arthur

R Yangarber

L Madoff

JP Linge

A Mawudeku

N Collier

JS Brownstein

G Thinus

N Lightfoot

ISSN 1471-2180

Article type Review

Received 09 July 2009

Revised 17 November 2009

Accepted 12 January 2010

© 2010 DM Hartley *et al.* ; licensee Emerging Health Threats Journal.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.5>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Landscape of International Event-based Biosurveillance

DM Hartley^{1*}, NP Nelson², R Walters³, R Arthur^{4**}, R Yangarber⁵, L Madoff⁶, JP Linge⁷, A Mawudeku⁸, N Collier⁹, JS Brownstein^{10,11}, G Thinus¹², and N Lightfoot¹³

1. Georgetown University School of Medicine, Imaging Science and Information Systems Center, 2115 Wisconsin Avenue NW, Suite 603, Washington, DC, 20057-1479 USA

(202) 687-9337 (voice) (202) 687-7231 (fax), hartley@isis.georgetown.edu

2. Georgetown University School of Medicine, Washington, DC, USA

3. Pacific Northwest National Laboratory, Richland, WA, USA

4. Centers for Disease Control and Prevention, Atlanta, GA, USA

5. University of Helsinki Department of Computer Science, Helsinki, Finland

6. University of Massachusetts Medical School, Worcester, MA, USA

7. Joint Research Centre, European Commission, Ispra, Italy

8. Public Health Agency of Canada, Ottawa, Ontario, Canada

9. National Institute of Informatics, Tokyo, Japan

10. Children's Hospital Boston, Boston, MA, USA

11. Harvard University Medical School, Boston, MA, USA

12. European Commission, Directorate for Public Health, Health Threats Unit, Luxembourg

13. Health Protection Agency, Central Office, London, United Kingdom

* Corresponding author

** The opinions expressed by the author do not necessarily reflect the opinions of the U.S. Department of Health and Human Services, the Public Health Service, or the Centers for Disease Control and Prevention.

ABSTRACT

Event-based biosurveillance is a scientific discipline in which diverse sources of data, many of which are available from the Internet, are characterized prospectively to provide information on infectious disease events. Biosurveillance complements traditional public health surveillance to provide both early warning of infectious disease events as well as situational awareness. The Global Health Security Action Group (GHSAG) of the Global Health Security Initiative is developing a biosurveillance capability that integrates and leverages component systems from member nations. This work discusses these biosurveillance systems and identifies needed future studies.

KEYWORDS

Surveillance, biosurveillance, infectious disease, epidemiology, biodefense, health security

INTRODUCTION

Far from being conquered by public health, vaccines, and antibiotics, infectious diseases continue to threaten humankind globally. There is a rich contemporary literature regarding the burden of endemic disease and epidemics of age-old threats, the emergence of newly discovered pathogens, drug resistance and the phenomenon of reemerging microbial threats¹⁻³. Additionally, biological terrorism remains a clear and present danger⁴. Beyond the personal impact on individuals suffering from infection, disease has societal impact: it can destabilize social institutions, populations, economies, and governments. For this reason, infectious disease is both a national and an international security issue^{5,6}.

The prevention and control of infectious diseases is therefore of extreme importance. World mobility rose significantly throughout the 20th Century and it continues to increase. Relative to past decades, people are traveling more, and travel times are dramatically shorter; at present it is possible to circumnavigate the globe in 36 hours via regularly scheduled commercial flights². More people, living species, and agricultural commodities are crossing borders than ever before, increasing the likelihood that pathogens circulating in one area will be trans-located to another area. One of the consequences of such global mobility is that disease prevention in any one area often depends on the effectiveness of surveillance, communication, and response control in other areas⁷.

Early warning of outbreaks may enable targeted quick intervention and control activities to take place. This was a motivation behind the 2005 revisions of the International Health Regulations (IHR)^{8,9}. The IHR-2005 provides an international legal framework for the early detection and reporting of, and response to, outbreaks of infectious disease. WHO member nations are obligated to develop and maintain surveillance, reporting, notification, verification, and response capabilities. Any nation with knowledge of a disease outbreak of international concern is obligated to report it to the WHO within 24 hours. The IHR-2005 are designed to ensure timely recognition of disease outbreaks of international public health significance and to promote effective containment before they spread.

Historically many epidemics have been reported through informal networks of health workers. Such networks should be timely, in order to assist in rapid detection, and sensitive, in order to detect potentially important outbreaks. As such, they may differ from traditional public health surveillance alluded to in the IHR, which often rely on classical epidemiologic

studies or clinical or laboratory data, the availability of which often lag the events they describe by days or months. This approach can also be less specific than traditional public health surveillance, though such trade-offs may be appropriate for a network designed to provide early warning.

Surveillance has been enhanced by the development of several novel approaches complementing traditional methods¹¹. Event-based biosurveillance is a new scientific discipline that uses information from the Internet whereby diverse streams of data are characterized prospectively to provide information on events affecting human health¹⁰. Indicator-based systems rely on routine collection of structured data such as syndromic surveillance and clinical activity monitoring, whereas these new event-based systems utilize unstructured data from media and other sources to detect anomalies that may indicate an emerging threat¹⁰. The potential of biosurveillance to contribute to global early warning of infectious disease and related threats, including chemical, biological, radiological, and nuclear (CBRN) agents, is becoming recognized¹². Researchers have developed prototype Internet-based systems to monitor and track the emergence of infectious disease and to evaluate the degree to which biosurveillance can provide early warning of outbreaks¹³.

Founded in 2001, the Global Health Security Initiative (GHSI) is an informal international partnership to strengthen health preparedness and response globally to CBRN terrorism threats and pandemic influenza¹⁴. Partners include Canada, European Union, France, Germany, Italy, Japan, Mexico, the United Kingdom and the United States with the World Health Organization (WHO) holding observer status. A Global Health Security Action Group (GHSAG) of senior officials from partner nations has been established by the GHSI to develop and implement concrete actions to improve global health security. The GHSI/GHSAG has established a number of working groups on areas such as smallpox, risk management and communication, chemical incidents, and pandemic influenza.

A GHSAG senior official meeting (in Ottawa, Canada, in June 2007) identified CBRN early warning as an area with great potential to support the efforts of GHSAG. A meeting of the Risk Management and Communications Working Group (RMCWG) (in Luxembourg, February, 2008) focused on identifying, within the context of CBRN hazards and risks, the capacities and input needs of existing IT systems working currently in the early detection of public health threats¹⁵. The RMCWG is currently making preliminary assessments of the

opportunities, with a focus on bioterrorism and diseases threatening public health. A follow-up meeting in Ispra, Italy explored in detail the tasks of each proposed work package in preparation for the Ninth Ministerial Meeting of the GHSI in Brussels, Belgium in early December 2008. In 2007-2008 the GHSAG made progress addressing key risks to global health security. This was accomplished through a variety of technical, scientific and policy networks and initiatives and stemmed from collective efforts and approaches in areas such as prevention, research, preparedness and response. In combination, the GHSAG event-based surveillance systems, which use the media as the primary source of information, form a unique part of the landscape of international biosurveillance.

METHODS

This review covers GHSAG-member biosurveillance systems, which constitute a major (though incomplete) fraction of similar capabilities available to the public health community at present. We elicited basic information from the respective system investigators in order to compare and contrast system capabilities and to illustrate the complementarities of the different approaches to event-based biosurveillance. Each biosurveillance system described in this study has been approved by the Institutional Review Board or corresponding authority at the respective institutions housing the systems.

SYSTEMS

Several systems originating from GHSAG member nations with a focus on biosurveillance or situation awareness are known at present and are described in this section. Table 1 provides a brief comparison of system traits and capabilities. The systems are listed alphabetically; no ranking should be inferred from the order of presentation.

Argus. Project Argus is a prototype bio-surveillance system designed to detect and track biological events that may threaten human, plant, and animal health globally¹⁶. The approach is based on monitoring social disruption evident in local, native-language media reports around the world. Argus employs analysts speaking approximately 40 languages to monitor a large number of media sources including traditional print and electronic media, internet-based newsletters, and blogs. It alerts users to events that may signal the initiation of outbreaks and shows trajectories of events that may require additional investigation. Bayesian analysis tools are utilized for article selection and alerting.

BioCaster (<http://www.biocaster.org>) is an experimental system for global health surveillance under development at the National Institute of Informatics in Japan and is a collaborative research project among five institutes in three countries¹⁷. The system is fully automated using Really Simple Syndication (RSS) feeds from over 1700 sources with no human analysts. Human analysis is assumed to take place down stream by the recipients of its output. BioCaster focuses on the Asia-Pacific region, posting approximately 90 articles per day in 3 languages (English, Japanese and Vietnamese) with plans for expansion to Thai, Chinese and other regional languages. Article capture and dissemination is done every hour. Until recently the primary sources are Google News, Yahoo! News, European Media Monitor, but the system is now expanding to take on sources from a commercial news aggregation company greatly increasing its coverage. BioCaster produces an ontology¹⁸ in 8 languages (Chinese, English, French, Japanese, Korean, Spanish, Thai and Vietnamese) that is openly available and is the basis for the Global Health Monitor¹⁹, an open access Web portal for displaying maps and graphs of health events to users. The ontology covers approximately 117 infectious diseases of humans and animals as well as 6 syndromes. Future objectives include extending language and health threat coverage.

Global Public Health Intelligence Network (GPHIN)

(<http://www.phac-aspc.gc.ca/gphin/index-eng.php>.) is the principal system used by WHO Alert and Response Operations for monitoring media articles²⁰. GPHIN was established in 1997 and is managed by the Public Health Agency of Canada's Centre for Emergency Preparedness and Response. GPHIN covers nine languages, Chinese Simplified and Traditional, Portuguese, Spanish, French, Russian, Arabic, and Farsi. GPHIN provides 24-hour operation 7 days per week coverage and applies human-based triage of information. Outputs are presented to GPHIN's user community (WHO, public health, intelligence, and law enforcement officials) as raw, machine-selected and translated articles. Machine selection of articles is based on keywords corresponding to the IHR. The online sources of GPHIN are drawn from sources available from Factiva and Al Bawaba, supplemented by automated and manual web crawling done by GPHIN analysts. Future objectives of GPHIN include integration of verification networks, addition of other news sources, additional languages, utilization of geographic and data visualization and incorporation of audio/visual feeds.

Health Emergency Disease Information System (HEDIS) (<http://hedis.jrc.it/>) Based in Italy, HEDIS is a situation awareness tool developed by the European Commission and aimed

primarily at crisis management. It supports the Health and Consumer Protection Directorate Genera (DG SANCO) and public health authorities in member states. There are approximately 300 users in Europe who utilize the system as an inter-organizational information-sharing platform to assist customers in dealing with an identified health threat. Users are Member States responsible for communicable disease and CBRN threats and risk communications. While HEDIS is not used for routine biosurveillance, in times of crisis it integrates biosurveillance information from many of the systems described in this study and makes it available to public health authorities.

HealthMap (<http://www.healthmap.org/about.php>) HealthMap is a multi-lingual, real-time disease outbreak detection, tracking and visualization system^{21,22}. Launched in fall 2006, the Web site collects over 300 reports per day in English, Spanish, French, Russian, Portuguese, Arabic and Chinese, from both general news media and public health sources around the world. Sources include Google News (in all seven languages) as well as other online news aggregators and informal sources, along with ProMED and WHO. The system also allows for user driven reporting through submission of URLs. Updated hourly, the system filters 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. Though the system is fully automated, a dedicated human analyst along with collaborators at US Centers for Disease Control and Prevention (CDC), the UK Health Protection Agency (HPA), WHO and ProMED examine reports each day to verify the accuracy of the system output. To date (as of 1 March 2009), the system has collected over 150,000 reports, covering 191 disease categories and over 200 countries and autonomous territories. With 1,000-150,000 visits per day (with top visitors from government, academic and public health agencies), HealthMap provides an overview of real-time information on emerging infectious diseases and has particular interest for public health officials and international travelers. Future system development is particularly focused on participatory surveillance, where users can contribute, edit and comment on disease intelligence as part of an online social network.

Medical Information System (MedISys) (<http://medusa.jrc.it/medisys/aboutMediSys.html>) MedISys is a fully automatic 24/7 public health surveillance system run and maintained by the Joint Research Centre (JRC) of the European Commission. The developer team

collaborates with the Health Threats Unit at the European Union Directorate General for Health and Consumer Affairs (DG SANCO) and the University of Helsinki (PULS system). MedISys covers infectious human and animal diseases, bioterrorism and CBRN threats reported in open-source news media. Approximately 90,000 articles from 5,000 news pages in 45 languages are screened. Currently, 26 languages are available via the Web portal, but news in all 45 languages is processed in pre-defined categories. Users can access world maps in which event locations are highlighted, graphs showing aggregated news counts by disease-location for an alerting category, graphs showing the most significant disease-location pairs for the last 24 hours, alerting statistics for regions of the world, filtering of news according to language, disease or location, and filtering by orthogonal categories such as “outbreaks”, “treatment”, “legislation”, and displaying of specific entities within the news article such as persons, organizations and search words. MedISys provides daily automated e-mail alerts to subscribers and offers a tool called Rapid News Service (RNS) in which users can manually select articles into pre-defined categories, comment on them, create formatted newsletters and distribute these to user defined groups. MedISys became online in August 2004 and is one of several JRC-developed media monitoring applications that process news gathered by the Europe Media Monitor (EMM, online since 2002). Therefore, future developments on EMM will also benefit MedISys²³⁻²⁵.

Program for Monitoring Emerging Diseases (ProMED) (<http://www.promedmail.org>) was established in 1994 and currently operates as a program of the International Society for Infectious Diseases with contributing corporate, foundation, and individual donor support^{26,27}. It is an un-automated, human-driven process, where over 40,000 freely subscribed members in over 160 countries submit reports of disease. The majority of these reports are media articles. Other sources include local observers, official reports and others. All reporting is screened by subject matter experts prior to posting (~7 reports issued per day). A total of 50,000 reports have been posted since project inception in the mid-1990s (10,000 of which are veterinary disease reports). ProMED has approximately 30 staff member subject matter experts, 5 regional programs and staff in 15 countries. Regional programs of ProMED include Latin America, the Mekong Basin, the East Africa Integrated Disease Surveillance Network (EAIDSNET), and ProMED-RUS (former Soviet Union). ProMED mail is available in English, Spanish, Portuguese, and Russian languages. Future objectives include French language reporting.

PULS (Pattern-based Understanding and Learning System)

(<http://puls.cs.helsinki.fi/medical/>) PULS is a project at the University of Helsinki, in collaboration with the European Commission's MedISys, and the European Centre for Disease Prevention and Control (ECDC). PULS traces its origins to the IFE-BIO Project, which aimed to analyze events reported in ProMED-Mail²⁸. As ProMED-Mail, PULS tracks human, animal and plant diseases, currently covering over 1,500 base terms, with a total of 2,500 variants. The focus in PULS is on the analysis of news texts for information extraction, aggregation, and visualization. PULS is fully automated with no human intervention. It uses MedISys as its main source, and employs natural language processing methods for analyzing the news stream to build a database of facts about the epidemiological events. The output of PULS is a spreadsheet-like view of the fact base, which is updated every 20 minutes. The base is also Google Earth-enabled. Linguistic coverage is primarily English, with a recent introduction of French-language analysis. The PULS average daily extraction rate varies from 300 entries during "normal" periods to over 1000 per day during times of heightened reporting, totaling about 300,000 entries to date. Future objectives include stronger multi-lingual support (with the addition of Spanish, Russian and Chinese), trend analysis, and data visualization.

DISCUSSION

Event-based biosurveillance possesses strengths and limitations that make it complementary to other experimental as well as traditional public health surveillance. Such systems may not always be timely, they may have limited specificity, and baseline thresholds for indicator detection may be difficult to quantify. While the systems described above are representative of the rapidly changing state of the art in event-based biosurveillance, important technological and methodological challenges remain²⁹. Prominent challenges include interoperability, interface customizability, scalability and event traceability. Integration of geospatial visualization, event mapping, modeling and trending tools are important for establishing metrics and baselines necessary for data interpretation and analysis. Additionally, expansion of the current biosurveillance capability via incorporation of emerging media such as video, audio, images, blogs, social networking sites, SMS (short message service) and others may be important.

While some qualitative aspects of recognizing important public health threats using event-based surveillance are evident, the value of diverse data sources must be quantified. Given the

diversity and richness of the Internet, and the availability of data and information from other sources (e.g., traditional public health; syndromic; and laboratory surveillance) of varying degrees of confidence and geographic coverage, how to quantify the payoff of including different sources in biosurveillance systems is unclear. Quantifying variation in source reporting standards as well as catchment (i.e., the regions from which a source collects data) and target population will be important for understanding the validity of biosurveillance system output. Metrics must be defined, and these metrics need to be generalizable across systems using different data and different approaches to analysis.

Standard guidelines for evaluating public health surveillance systems may not be wholly appropriate for evaluating event-based biosurveillance systems³⁰. Techniques for evaluating system performance are needed and standardized metrics quantifying the performance of distinct biosurveillance systems must be developed. Such metrics are also needed if end users are to be able to understand the performance of a given system, or an aggregation of systems. Similarly, analytic methods for assessing and quantifying the value added by biosurveillance to other approaches to surveillance and situational awareness must be developed.

Efficient and meaningful ways of communicating complex biosurveillance data must be identified. Because they are tailored to meet the needs of their specific user communities, current systems display and present the results of biosurveillance differently. How to best present results to the broader user community, which includes researchers as well as public health workers and decision makers, is unclear. Many unknowns remain, including identifying the most appropriate interactive visual interfaces; best practices regarding techniques for synthesizing biosurveillance data visually; and how to present dynamic, ambiguous, and potentially conflicting information to consumers of biosurveillance.

Real time situational awareness of emerging biological threats is needed in today's dynamic world. However, if such an approach to public health response is to be viable, a capability must exist to detect evidence of outbreak activity at the earliest stages and monitor related information as it evolves. We are unaware of published studies investigating the timeliness of event-based biosurveillance using Internet sources relative to traditional approaches to public health surveillance. In order to maximize the likelihood of early detection, such a capability should be composed of discrete components acting in concert. At one end of the alerting spectrum, biosurveillance systems that provide indications and warning (I&W) of potential

infectious disease events are needed. These I&W components would provide the first tip of a potential event or risk of a future event. Necessarily, information provided by systems at this end of the spectrum would have limited confidence and their output would need to be refined and better characterized by other components in the alerting spectrum. Toward the middle of the spectrum would be systems that more-directly measure infectious disease activity, e.g., syndromic surveillance systems. At the opposite end of the spectrum would be traditional formal clinical and laboratory-based public health surveillance.

At the biosurveillance end of the spectrum, there is considerable variation in system capability, data analyzed, and products disseminated, pointing to the need for integration. A recent meeting of GHSAG participants (Luxembourg, 2008) highlighted the need for “cooperation at all levels, between systems, between systems and users, and users amongst themselves. Such cooperation should be considered at the level of the collection of data, at the level of data analysis of the data available and the subsequent sharing of the relevant information through a common restricted platform.”³¹

While such a capability does not yet exist, similarities and differences among the systems described above suggest that combining these approaches into a single system can provide a powerful biosurveillance resource. The GHSAG is developing such a prototype biosurveillance “system of systems”; it is anticipated that, with appropriate communication and data sharing protocols, technical barriers to integrating existing global and regional biosurveillance systems can be overcome. Partially because each of the individual systems examined here have different missions and approaches, they complement one another. This complementarity will be demonstrated in the GHSAG pilot integration project.

REFERENCES

1. Oldstone MBA. Viruses, plagues, and history. New York: Oxford University Press. 1998.
2. Institute of Medicine, Microbial Threats to Health: Emergence, Detection and Response. Washington, DC: National Academy Press. 2003.
3. Mathers C, Boerma T, Ma Fat D. Global Burden of Disease 2004 Update. World Health Organization Press: Geneva, Switzerland. 2008
4. Institute of Medicine, Biological Threats and Terrorism: Assessing the Science and Response Capabilities. Washington, DC: National Academy Press. 2002.
5. Calain P. Exploring the international arena of global public health surveillance. *Health Policy Plan.* 2007 Jan;22(1):2-12.
6. King NB. 2002. Security, disease, commerce: ideologies of postcolonial global health. *Social Studies of Science* 32:763–89.
7. Rodier G, Greenspan AL, Hughes JM, Heymann DL. Global public health security. *Emerg Infect Dis.* 2007 Oct;13(10):1447-52
8. Baker MG, Forsyth AM. The new International Health Regulations: a revolutionary change in global health security. *N Z Med J.* 2007 Dec 14;120(1267):U2872.
9. Wilson K, McDougall C, Forster A. The responsibility of healthcare institutions to protect global health security. *Healthc Q.* 2008;12(1):56-60.
10. Walters R, Harlan P, Nelson N, Hartley D. Data Sources for Biosurveillance. In: Voeller J, editor. *Wiley Handbook of Science and Technology for Homeland Security.* New York: John Wiley & Sons, Inc. Forthcoming 2009.

11. Amato-Gauci A, Ammon A. The surveillance of communicable diseases in the European Union--a long-term strategy (2008-2013). *Euro Surveill.* 2008 Jun 26;13(26). pii: 18912.
12. Institute of Medicine. *Global Infectious Disease Surveillance and Detection: Assessing the Challenges. Workshop Summary.* Washington, DC: National Academy Press. 2007.
13. Wilson K, Brownstein JS. Early detection of disease outbreaks using the Internet. *CMAJ.* 2009 Apr 14;180(8):829-31
14. Global Health Security Initiative. GHSI background [Internet]. [place unknown]: Overview: Global Health Security Initiative (GHSI) [Cited 4 July 2009]. Available from: [<http://www.ghsi.ca/english/background.asp>].
15. European Commission. GHSAG Working Group on risk management and communication, Meeting on Early warning for CBRN threats, Luxembourg – 26 to 28 February 2008, Short summary of the meeting. [Location unknown]: European Commission, Health & Consumer Protection Directorate-General, Directorate C - Public Health and Risk Assessment, C3 - Health Threats. [Cited 4 July 2009]. Available from: [http://www.episouth.org/doc/r_documents/GHSAG_short_report_for_PHEA_web.pdf].
16. Reilly AR, Iarocci EA, Jung CM, Hartley DM, Nelson NP. Indications and Warning of Pandemic Influenza Compared to Seasonal Influenza. *Advances in Disease Surveillance* 2008;5:190
17. Collier N, Doan S, Kawazeo A, Goodwin R, Conway M, Tateno Y, Hung-Ngo Q, Dien D, Kawtrakul A, Takeuchi K, Shigematsu M, Taniguchi K (2008), "BioCaster: detecting public health rumors with a Web-based text mining system", *Bioinformatics*, Oxford University Press, DOI: 10.1093/bioinformatics/btn534.
18. Collier N, Kawazoe A, Jin L, Shigematsu M, Dien D, Barrero R, Takeuchi K, Kawtrakul A. A multilingual ontology for infectious disease outbreak surveillance: rationale, design and challenges. *Journal of Language Resources and Evaluation*, Springer, 2007 vol. 40, no. 3-4, pp.405-413. DOI information: 10.1007/s10579-007-9019-7

19. Doan S, Hung-Ngo Q, Kawazoe A, Collier N. Global Health Monitor - A Web-based System for Detecting and Mapping Infectious Diseases. Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP), Companion Volume, 2008 pp. 951-956
20. Mykhalovskiy E, Weir L. The Global Public Health Intelligence Network and early warning outbreak detection: a Canadian contribution to global public health. *Can J Public Health*. 2006 Jan-Feb;97(1):42-4.
21. 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
22. Brownstein JS, Freifeld CC, Reis BY, Mandl KD. Surveillance Sans Frontières: Internet-based emerging infectious disease intelligence and the HealthMap project. *PLoS Med*. 2008 Jul 8;5(7):e151.
23. Steinberger R., Flavio F., van der Goot, E., Best, C., von Etter, P. and Yangarber, R. (2008), "Text Mining from the Web for Medical Intelligence", In: Françoise Fogelman-Soulié, Domenico Perrotta, Jakub Piskorski & Ralf Steinberger (eds.): *Mining Massive Data Sets for Security*. IOS Press, Amsterdam, The Netherlands.
24. Best, C., van der Goot, E., Blackler, K., Garcia, T. and Hornby, D. (2005), "Europe Media Monitor – system description", EUR, Technical Report 22173 EN.
25. Linge JP, Steinberger R, Weber TP, Yangarber R, van der Goot E, Al Khudhairy DH, Stilianakis NI. Internet surveillance systems for early alerting of health threats. *Euro Surveill*. 2009 Apr 2;14(13)
26. Cowen P, Garland T, Hugh-Jones ME, Shimshony A, Handysides S, Kaye D, Madoff LC, Pollack MP, Woodall J. Evaluation of ProMED-mail as an electronic early warning system for emerging animal diseases: 1996 to 2004. *J Am Vet Med Assoc*. 2006 Oct 1;229(7):1090-9.
27. Madoff LC. ProMED-mail: an early warning system for emerging diseases. *Clin Infect*

Dis. 2004 Jul 15;39(2):227-32.

28. Grishman R, Huttunen S, Yangarber R. Information Extraction for Enhanced Access to Disease Outbreak Reports. *J Biomed Inform.* 2002 Aug;35(4):236-46.
29. Brownstein JS, Freifeld CC, Reis BY, Mandl KD, (2007) HealthMap: Internet-based emerging infectious disease intelligence. In: Institute of Medicine, editor *Infectious disease surveillance and detection: Assessing the challenges—Finding solutions* Available http://books.nap.edu/openbook.php?record_id=11996&page=122. Accessed 6 June 2008.
30. Buehler JW, Hopkins RS, Overhage JM, Sosin DM, Tong V; CDC Working Group. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC Working Group. *MMWR Recomm Rep.* 2004 May 7;53(RR-5):1-11
31. European Centre for Prevention and Disease Control. Framework for a strategy for infectious disease surveillance in Europe. 2006. Online at ecdc.europa.eu/documents/pdf/Framework%20surveillance%20strategy%20in%20Europe.pdf . (accessed 14 March 2009)

FIGURES AND TABLES

Table 1: Comparison of System Traits and Capabilities

System Name	Argus	Biocaster	GPHIN	HEDIS
Sponsoring Agency	US Government	National Institute of Informatics in Japan	Canadian government	European Commission
Access Policy	Limited	Open to public	Fee-based	Restricted
Posts Per Day (Approx.)	120-170	90	2000	NA
Staff	45	5	9 daytime 6 nighttime	5 Ispra 3 Luxembourg
Geographical Coverage	World-wide, not US	Asia-Pacific region	World-wide	World-wide
Languages	40	3	9	NA
Update Frequency	Every 1-15 minutes	Every hour	Every 20 minutes	News: continuous Documents: when uploaded
Covered Topics	Human, animal, plant diseases, enviro-climatic indicators	Human, animal infectious diseases	Human, animal, plant diseases; chem/rad events; 'unsafe products' and nat. disasters	Health Threats
Information Disseminated	Watch-board and alerts	Watch-board	Watch-board; approx. 1 – 4 'alert' emails delivered to subscribers daily	Watch-board
Staging or Scoring System	Yes, 7-stage social disruption scale at the event level; low, medium, high level of importance at the country level	No	"Relevancy Scoring" system and IHR related decision tree is applied that provide email "Alerts"	No
Direct Indicators and Warning	Public health	Public Health	Public health	Public Health
Indirect Indicators and Warning	Public health response, meteorological data, other government reaction, business/organization changes, other social behavior	No	Public health response, other government reaction, meteorological data	No

Table 1: Comparison of System Traits and Capabilities (Continued)

System Name	HealthMap	MedISys	ProMed Mail	Puls
Sponsoring Agency	Children's Hospital Boston, Harvard University soft monies	European Commission	ISID, soft monies	University of Helsinki and European Commission's Joint Research Center
Access Policy	Open to public	Open to public	Open to public	Public and restricted views
Posts Per Day (Approx.)	30-300	Not applicable	7	300
Staff	6	20 (for both MedISys and Europe Media Monitor)	30	7
Geographical Coverage	World-wide	World-wide	World-wide	World-wide
Languages	7	26	4	2
Update Frequency	Every hour	Every 10 minutes	Variable	Every 20 minutes
Covered Topics	Human, animal, and plant diseases	Human and animal diseases, and very loosely defined 'health' topics	Human, animal, plant diseases	Human, animal, plant diseases
Information Disseminated	Watch-board, automated Email alerts, RSS, mobile phone, Twitter,	Watch-board; email 'alerts' auto-delivered to subscribers	Watch-board	Spreadsheet-like tables
Staging or Scoring System	An 11-staged "Heat Index" system	Alerts at low, medium, and high levels are assigned to reports	No	Relevance scoring: importance to user (1-5); system confidence score
Direct Indicators and Warning	Public health, clinical, laboratory	Public health	Public health, clinical, laboratory, veterinary	Public health
Indirect Indicators and Warning	Public health, Environmental disasters, Conflicts	No	Public health	No