Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

A systematic literature review on even-based public health surveillance systems

Hamburg University of Applied Sciences

Faculty of Life Sciences

Master of Health Sciences

Date of Submission:

14.09.2011

Presented by:

First Name: Tumacha

Last Name: Agheneza

First Supervisor:

Prof. Dr. Ralph Reintjes
University of Applied Sciences Hamburg
Faculty of Life Sciences

Second Supervisor:

Dr. Tim Eckmanns
Robert Koch Institute
Dep't of Infectious Disease Epidemiology

Statutory Declaration

I hereby declare that I have developed and written the enclosed Master Thesis completely by myself without any external unauthorised help. It has neither been presented to any institution for evaluation nor previously published in its entirety or in parts. Any parts, words or ideas, of the thesis, however limited, and including tables, graphs, etc., which are quoted from or based on other sources, have been acknowledged as such without exception.

I agree that a copy of this thesis may be made available in the Library of University of Applied Sciences Hamburg.

I agree that the abstract of this thesis may be placed on the website of University of Applied Sciences Hamburg.

Tumacha Agheneza

Berlin, 7th September 2011

Abstract

Introduction: Public health authorities increasingly rely on systems that perform epidemic intelligence constituting of indicator-based and event-based components in order to gather a comprehensive picture of potential epidemic threats. The indicator-based component collects and analysis *structured* data systematically while the event-based component collects and analysis' *unstructured* data from websites, social media, discussion groups, blogs etc. The event-based component now allow us to shortcut traditional reporting mechanisms that travel through the various levels of public health administration and thus allow us to detect disease outbreaks earlier with reduced cost and increased reporting transparency. We want to find out all the event-based surveillance systems exist, where they are based and which systems have been evaluated.

Methods: We searched for scientific literatures in Pubmed, Scopus and Scirus retrieving about 39,000 articles. Going through our inclusion exclusion criteria, we ended up with 32 articles.

Results: 13 event-based systems were identified, and with 10 of the 13 systems evaluated. N. America is the leading continent with about 77 % of the event-based systems followed by Europe and lastly Asia. No system was found in Africa, Australia and S. America.

Conclusion: With Africa, Asia and S. America being the 3 most affected continents with emerging infectious diseases possessing little or no event-based systems to monitor their epidemic threats, there is need for these continents to take the advantage in the advancement of modern information technology to set up their event-based systems which could be relatively cheap but effective to complement their indicator-based systems. This review may therefore provide the necessary background to public health officials in developing new event-based systems or enhancing their indicator-based work.

Zusammenfassung

Einleitung: Gesundheitsbehörden benötigen zunehmend Surveillance-Systeme, die ein umfassendes Bild der möglichen epidemiologischen Bedrohungen bieten. Traditionelle Indikator-basierte Surveillance beinhaltet zwar eine systematische Analyse von Meldedaten kann aber mit zusätzlichen Komponenten verstärkt werden, z. B. durch so genannte event-basierten Aktivitäten, die unstrukturierter Daten aus Websites, „Social Media", Diskussionsgruppen, Blogs sammeln und zu Verfügung stellen. Ziel dieser Arbeit war es, event-basierten Surveillance-Systeme zu erforschen, in welchem Land sie angesiedelt sind und welche Systeme bisher bereits evaluiert sind.

Methoden: Eine struktuierte Literatursuche wurde durchgeführt. Als Quellen wurden Pubmed, Scopus und Scirus genutzt und rund 39.000 Artikeln abgerufen.

Ergebnisse: Dreizehn event-basierte Systeme wurden identifiziert und 10 der 13 Systeme evaluiert. Nord America hat mit über 77% die meisten event-basierten Systeme. Es sind keine Systeme in Afrika, Australien und Südamerika vorhanden.

Fazit: Obwohl Afrika, Asien und Südamerika die 3 stärksten betroffenen Kontinenten für aufkommende Infektionskrankheiten sind, sind in diesen Länder keine event-basierten Systeme vorhanden. Es ist notwendig, dass Afrika, Asien und Südamerika die Vorteile neuer Informationstechnologien für event-basierte Systeme nutzt, diese sind kostengünstige Systeme zur Unterstützung von indikator-basierten Systemen. Diese Untersuchung kann daher den notwendigen Hintergrund für öffentliche Gesundheitsdienste bieten, damit diese neue event-basierten Systeme entwickeln oder vorhandene indikator-basierte Systeme verbessern.

Acknowledgements

I am heartily thankful to my second supervisor Dr. Tim Eckmann and Mr. Edward Velasco both at the Robert-Koch Institute (RKI), whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the subject. Additionally, I want to thank Mr. Edward Velasco for the good advice, support and friendship that has been invaluable on both academic and personal level for which, I am extremely grateful.

Above all, I would like to thank my principal supervisor Prof. Dr. Ralf Reintjes for his personal support and great patience in correcting my work at all times not to mention his advice and unsurpassed knowledge on Epidemiology and surveillance.

I would like to acknowledge Evangelische Studentinnen- und Studentengemeinde (ESG) Hamburg, Soziales & Internationales Studierendenwerk Hamburg as well as RKI for their financial assistance during my internship and thesis phases while studying in Berlin.

Last, but by no means least, I thank all the co-workers of Surveillance Unit (Concentration 32) at RKI for their support and encouragement throughout. They were particularly good around me.

For any errors or inadequacies that may remain in this work, of course, the responsibility is entirely my own.

Tumacha Agheneza

Table of Contents

List of Tables

List of Figures

List of Abbreviations Used

	Acronym	Meaning
1	API	Application Programming Interface
2	BCO	BioCaster ontology
3	CDC	Center for Disease Control and Prevention
4	CEPR	Centre for Emergency Preparedness and Response
5	EC	European Commission
6	ECDC	European Centre for Disease Prevention and Control
7	EEA	European Economic Area
8	EI	Epidemic Intelligence
9	EISN	European Influenza Surveillance Network
10	EMM	Europe Media Monitor
11	EpiSPIDER	Semantic Processing and Integration of Distributed Electronic Resources for Epidemiology
12	EU	European Union
13	EWRS	Early Warning and Response System
14	FAO	Food and Agricultural Organization
15	FAS	Federation of American Scientists
16	FDA	Food and Drug Administration
17	GDACS	Global Disaster Alert Coordinating System
18	GIS	Geographic Information System
19	GOARN	Global Outbreak and Alert Response Network
20	$GODS^N$	Global News Driven Disease Outbreak and Surveillance
21	GPHIN	Global Public Health Intelligence Network
22	IE	Information Extraction
23	ILI	influenza-like-illness
24	InSTEDD	International System for Total Early Disease Detection
25	IR	Information Retrieval
26	ISID	International Society for Infectious Diseases
27	ITIC	Intelligence Technology Innovation Center
28	JRC	Joint Research Centre
29	JSON	JavaScript Object Notation
30	KML	Keyhole Markup Language
31	MedISys	Medical Information System
32	MiTAP	MITRE Text and Audio Processing System
33	NGO	Non-governmental organization.
34	NOAA	National Oceanic and Atmospheric Administration
35	OIE	World Organisation for Animal Health
36	OWL	Web Ontology Language
37	PH	Public Health
38	PMM	ProMED-mail
39	ProMED	Program for Monitoring Emerging Diseases
40	PULS	Pattern-based Understanding and Learning System
41	RKI	Robert-Koch Institute

	Acronym	Meaning
42	RNS	Rapid News Service
43	RSS	Really Simple Syndication
44	SARS	Severe Acute Respiratory Syndrome
45	SMS	Short Message Service
46	TV	Television
47	UK	United Kingdom
48	UN	United Nations
49	UNICEF	United Nations International Children's Emergency Fund
50	US	United States
51	USA	United States of America
52	USDA	United States Department of Agriculture
53	WHO	World Health Organization

1. Introduction

Public health surveillance is the ongoing, systematic collection, analysis, interpretation, and dissemination of data regarding a health-related event for use in public health action to reduce morbidity and mortality and to improve health.[1–3] The rapid identification of an infectious disease outbreak is critical, both for effective initiation of public health intervention measures and timely alerting of government agencies and the general public but the surveillance capacity for such detection can be costly, and many countries lack the public health infrastructure to identify outbreaks at their earliest stages.[4] Current public concern over the spread of infectious diseases has underscored the importance of health surveillance systems to quickly detect disease outbreaks.[5]

Surveillance systems, therefore, consist of routine data collection and data analyses, followed by a response when required. It is this element of decision and timely response based on interpretation of the data that makes surveillance different from monitoring, making it more than just a system for event detection as illustrated in Figure 1.

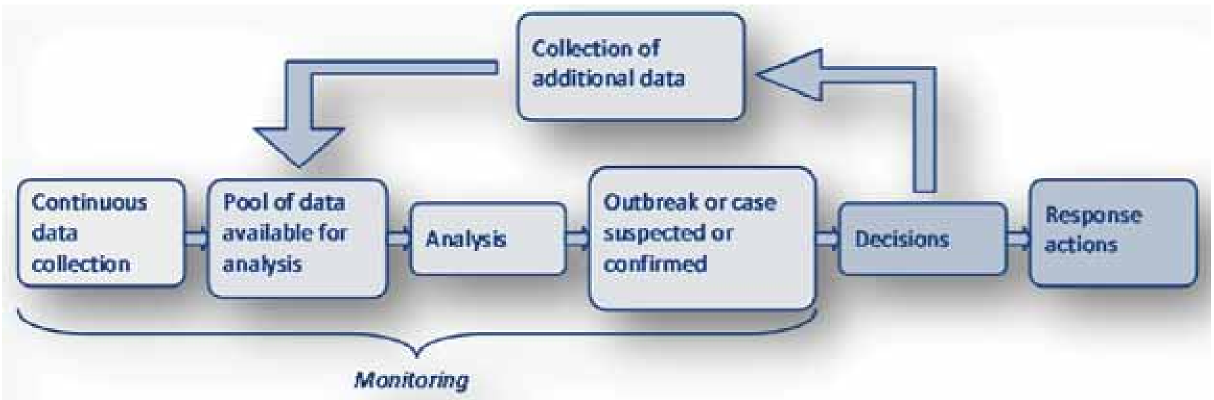

Figure 1 Structure of a surveillance system. Data collection to detection of an outbreak or case consists of monitoring; the addition of timely decisions and response makes the system a surveillance system.[6]

Surveillance is ‘action-oriented’, wherein ‘real-time’ decisions are linked to current findings. Therefore, surveillance includes timely response to the data. Monitoring, on the other hand, does not necessarily include a timely response, but sacrifices timeliness for accuracy.[7] The differences between surveillance and monitoring have been summarized in Table 1.

Table 1 Comparing between surveillance and monitoring systems

Surveillance	Monitoring
Active	Passive
Continuous	Episodic
Timely dissemination of results to decision makers for rapid intervention	Sacrifices time for accuracy

1.1 Epidemic Intelligence (EI)

Public health authorities increasingly rely on systems that perform epidemic intelligence (EI) in order to gather a comprehensive picture of potential epidemic threats.[8] EI encompasses all activities related to early identification of potential health hazards, their verification, assessment and investigation in order to recommend public health control measures. It makes use of information that originates from official sources such as national public health surveillance systems as well as from informal sources such as electronic media and web-based information tools. It integrates both an indicator-based and an event-based component [9] as illustrated in Figure 2.

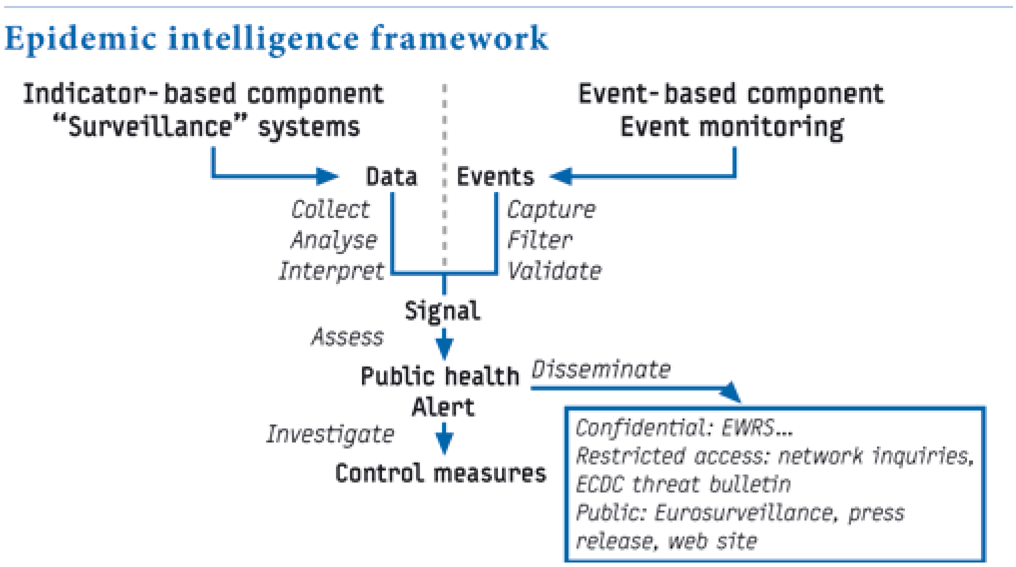

Figure 2 Existing systems for supporting epidemic intelligence can be grouped into two separate surveillance systems: *indicator-based* and *event-based* surveillance systems [9]

1.1.1 Indicator-based component

This component collects and analysis **structured** data systematically. Indicator-based surveillance is the traditional approach to the surveillance of communicable disease which consists of routinely collecting data about the occurrence of predefined diseases, specific pathogens, syndromes or conditions from health-care providers. This notification process relies on standard case definitions for surveillance to ensure a uniform approach to reporting by all clinicians and laboratories and to improve the comparability of the data and reports across healthcare services. The notifications are then routinely compiled and analysed to produce indicators that could suggest the existence of a threat or a problem that needs addressing. This indicator-based approach has proved to be very effective in monitoring threats related to known risks and then in ensuring the prompt implementation of public health measures.

Even though this traditional approach remains the backbone of public health surveillance for communicable diseases, it has proven to be less effective in ensuring prompt recognition of emerging problems. Several further approaches seek to complement traditional surveillance in order to enhance its ability to detect public health threats.[10]

1.1.2 Event-based component

This component collects and analysis **unstructured** data. Event-based surveillance is a novel approach which takes advantage of the availability of advanced information technology by scanning sources such as the Internet and media continuously to detect information that may lead to the recognition of emerging threats. This event-based surveillance[9] approach was introduced to complement effectively the indicator-based surveillance approach. It uses unstructured data, which then needs to be studied and verified and hence, cannot be summarized as an indicator.[10]

The strengths of event-based surveillance systems are based on the (unconscious) reporting of relevant information which helps to detect:

- rare and new events not specifically included in indicator-based surveillance, and
- events that occur in populations which do not access health care through formal channels or where indicator-based systems do not exist.

Event-based surveillance can also be distinguished from indicator based surveillance based on the definitions as shown in Table 2 (page 4).

Table 2 Differences between indicator-based and systems event-based

Indicator based	Event-based
'Indicator-based' refers to **structured** data collected through routine surveillance systems[11]	'Event-based' refers to **unstructured** data gathered from sources of intelligence of any nature
Routine reporting of cases of disease, including • Notificable disease surveillance system • Sentinel surveillance • Laboratory based surveillance[12] • Syndromic surveillance[13]	Rapid detection, reporting, confirmation, assessment of public health events including • Clusters of disease. • Rumours of unexplained deaths

The characteristics of indicator-based and event-based systems can been summarized in Table 3.

Table 3 Characteristics of indicator-based and event-based systems (a pro is marked by "+", a contra marked by "-")[14]

	Indicator-based systems	+/-	+/-	Event-based systems
Timeliness of data input	• Immediate / weekly / monthly • Possible delay between identification and notification	-	+	• Information might be available immediately after the occurrence of an event • Real time
Reporting structure	• Clearly defined • Reporting forms • Reporting dates • Teams to analyse the data in regular intervals	+	-	• Predefined or not predefined structure • Reporting forms flexible for qualitative and quantitative data • At any time • Teams confirm events and prepare the response
Trigger for follow-up or action	• Crossing a pre-defined threshold leads to an in-depth analysis and further information gathering	+	+	• A confirmed event or even a rumour leads to further information gathering, verification
Timeliness of detection	• Depends mainly on two time intervals. First the time from the onset of the disease until a diagnosis is available that fulfils the case definition. Second the time for reporting through the stages of a hierarchical reporting structure	-	+	• Depends on the time from the onset of the disease until the first mention occurs, which might be before diagnostic confirmation is available
Thresholds for signal generation	• Statistical methods are used for cluster detection	+	-	• Signals are differentially generated, i.e. human indexing in ProMed-mail, but rarely with statistical methods. The relevance of the signal is derived from the fact that the information about the presumed event is published.

1.2 Research questions

- Which are all the event-based surveillance systems that exist?
- In which countries or continents are they based?
- How do they collect, analyse, interpret and disseminate information (to decision makers)?
- Which systems have been evaluated?
- Which systems have been shown to be effective?

1.3 Purpose of Study

The purpose of this systematic literature review is to find out which event-based surveillance systems exist till date and which systems have been evaluated. The need for seeking evaluation of public health surveillance systems is to ensure that problems of public health importance are being monitored efficiently and effectively.[3]

1.4 Importance of Study

Till now, a systematic literature review on event-based surveillance systems has not been completed. This research is worth doing because:

1. It will provide a useful report that is up to date with what is current in the field of event-based surveillance.
2. It will help to generate criteria for evaluating such existing systems (Table 20; Appendix 1).
3. It will provide background for developing new systems or enhancing indicator-based work.

1.5 Previous work and scope of this study

We found two reviews that were closely related to this study. The first study was the work of Bravata et al.,[15] which was a systematic review on surveillance systems for early detection of bioterrorism-related diseases. Their purpose was to critically evaluate the potential utility of existing surveillance systems for illnesses and syndromes related to bioterrorism. For the Methods, their data sources were databases of peer-reviewed articles (for example, MEDLINE for articles published from January 1985 to April 2002) and web sites of relevant government and nongovernment agencies. They reviewed 17,510 article citations and 8088 government and nongovernmental web sites. In their result, the authors included 115 systems that collect various surveillance reports, including 9 syndromic surveillance systems, 20 systems collecting bioterrorism detector data, 13 systems collecting influenza-related data, and 23 systems collecting laboratory and antimicrobial resistance data. Only the systems collecting syndromic surveillance data and bioterrorism detection system data were designed, at least in part, for bioterrorism preparedness applications. Syndromic surveillance systems have been deployed for both event-based and continuous bioterrorism surveillance and only 3 systems have had both sensitivity and specificity evaluated.

The second review was the work of Vrbova et al.,[6] where the author conducted a systematic review of surveillance systems for emerging zoonotic diseases. Their purpose was to synthesize available evidence for public health practitioners making decisions in the event of an emerging zoonosis, by finding public health surveillance initiatives for emerging zoonoses, and seeing what criteria have been used to evaluate these systems. For the Methods they reviewed peer-reviewed articles from databases such as MEDLINE, EMBASE, AGRICOLA, and a few others that described and/or evaluated surveillance systems for emerging zoonotic diseases between 1992 and 2006. Their results revealed that out of the 221 systems identified only 17 systems were evaluated.

Till now, a systematic literature review on event-based surveillance systems has not been completed. Since a substantial amount of work has been done on bioterrorism related systems[15], the scope of this work will not include bioterrorism related studies, unless the study itself is describing an event-based surveillance system.

In this study, we conducted a systematic review focusing only on event-based surveillance systems. In the following sections we presented our Method where the search methodology was to extract articles in the time period from 1990 - 2011, from three databases; Pubmed, Scopus and Scirus and retained articles based on our inclusion and exclusion criteria (see section 2.3 on page 8). Next, we presented the results followed by the discussion and recommendations. We then summarized our finding in the conclusion and lastly, we included the limitations of this work.

2. Methods

2.1 Preliminary Search

A previous literature search on this topic was conducted in which the first step was to generate a list of keywords.

The keyword search was not the main step for a literature search but just the initial. The keywords were determined by searching a few published articles to see what keywords were used by those authors in this topic.

The keywords that appeared in most of those literatures were chosen to be the most potential ones. After this initial search, a further backward reference and author search helped to produce more keywords. After reviewing those articles retrieved and a few articles gotten by referral from Robert Koch Institute (RKI), a set of inclusion and exclusions criteria were determined. A forward search was performed using Pubmed, Scopus, Scirus, Trip database and a few other databases. After reviewing them by the inclusion and exclusion criteria as well as the full texts, a total of 130 articles were retained. This set of articles were included in the second literature search conducted during the Thesis phase and were identified as "preliminary search" as illustrated in Figure 3 (page10).

2.2 Study Selection

The set of keywords that were generated from the preliminary search have been reviewed by a committee made-up of three persons (the author and 2 epidemiologists at RKI) and a new set of keywords (Table 4) were adopted based upon consensus.

Table 4 List of keywords

		OR
Surveillance	AND	early detection
		early warning
		electronic
		electronic media
		epidemic intelligence
		event-based
		internet-based
		media
		media based
		news

		OR
		online media
		outbreak
		reporting system
		social media
		syndromic surveillance
		warning system
		web-based

In order to widen the scope of our search and to increase the chances of retrieving more articles, 3 databases were chosen, which were; Pubmed, Scopus and Scirus.

PubMed was chosen because it is a service of the National Library of Medicine which provides access to over 12 million MEDLINE citations and additional life science journals. It includes links to many sites providing full text articles and other related resources.[16] Scopus is an integrated part of SciVerse, thus SciVerse Scopus is the world's largest abstract and citation database of peer-reviewed literature and quality web sources with smart tools to track analyze and visualize research. SciVerse Scopus is the most direct way to find relevant content.[17] Scirus searches only scientific information and it is the most comprehensive scientific research tool on the web. With over 410 million scientific items indexed at last count, it allows researchers to search for not only journal content but also scientists' homepages, courseware, pre-print server material, patents and institutional repository and website information.[18]

2.3 Inclusion and Exclusion Criteria

This set of inclusion and exclusion criteria, is a revised version of the inclusion and exclusion criteria from the preliminary literature search. A committee of 3 persons (the author and 2 epidemiologists at RKI) agreed on this inclusion/exclusion based on consensus:

2.3.1 The inclusion criteria

- Infectious diseases
- Surveillance
- Outbreak

- Describes an event-based surveillance system
- Only human systems
- Only English articles

2.3.2 The exclusion criteria

- Bioterrorism
- Technical aspects
- Security (e.g. video surveillance)
- Sentinel surveillance
- Surveillance whose primary aim is not based on early detection of outbreak
- No abstract available.

2.4 Search Methodology

Our search strategy was based on a broad search. Even though a broad search resulted in the inclusion of many papers which were later deemed irrelevant, it reduced potential biases that a narrower search would have produced, in other words, it increased the retrieval of relevant studies. For example, the aim of this study was to synthesize only human event-based surveillance systems, but the option to retrieve only human articles in Pubmed was not activated (found in the option “limit”) during the search so as to minimize biases caused by narrower searches. Irrelevant studies were instead eliminated during the review of articles by inclusion/exclusion criteria.

Given the fact that each database was different due to the difference in subject thesauri or subject terminology, the exact search strategy was unique for each database. The search was conducted as illustrated in Figure 3.

Systematic literature review on event-based surveillance systems

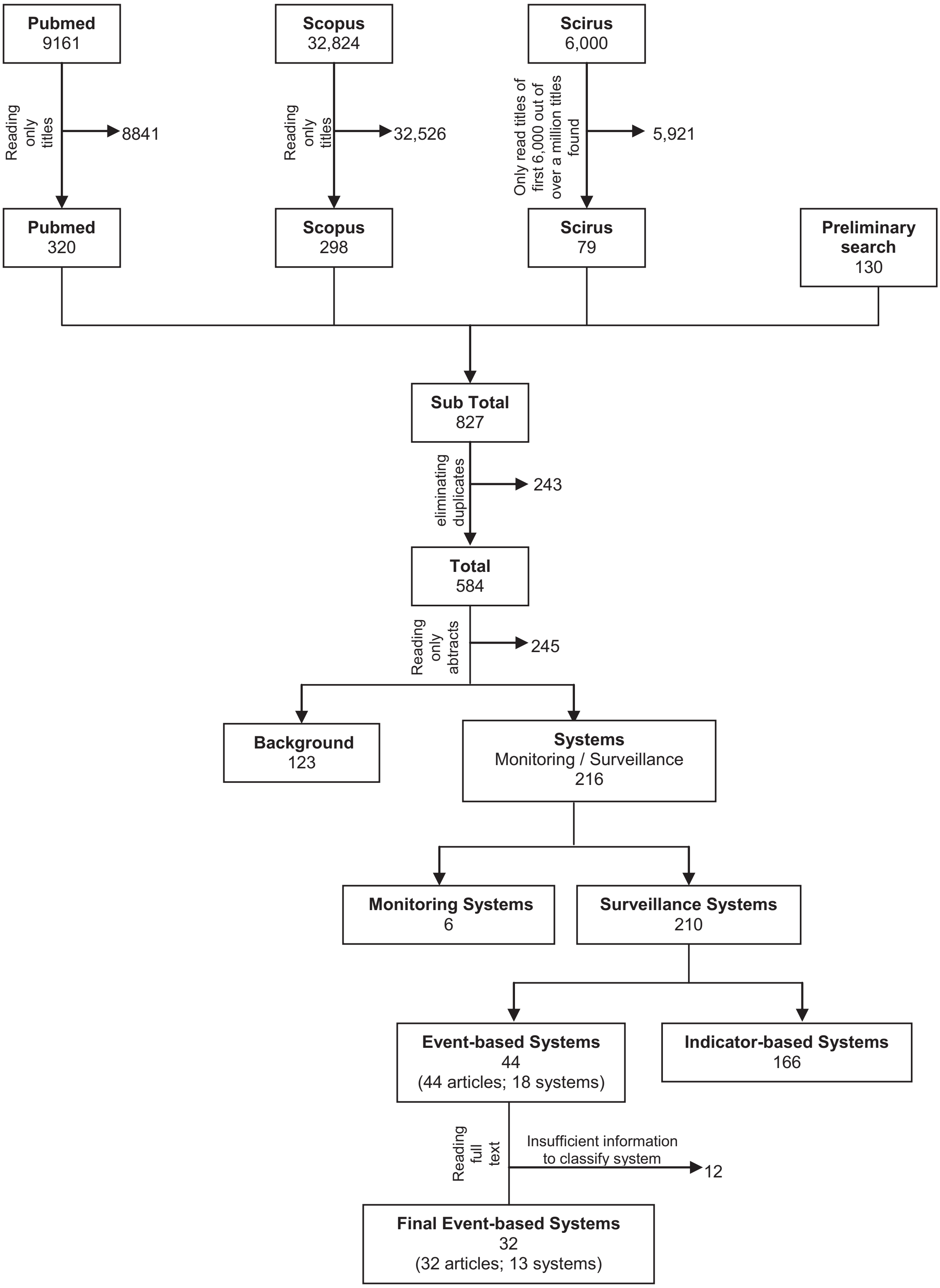

Figure 3 Flow chart of search procedure illustrating inclusion and exclusion processes.

2.4.1 Pubmed

Pubmed was the first database for which the search was conducted. The search was conducted on 07.04.2011. The set of keywords were combined as follows:

Surveillance[Title/Abstract]) *AND* (infectious disease[Title/Abstract] *OR* early detection[Title/Abstract] *OR* early warning[Title/Abstract] *OR* electronic[Title/Abstract] *OR* epidemic intelligence [Title/Abstract] *OR* event-based [Title/Abstract] *OR* internet-based[Title/Abstract] *OR* media[Title/Abstract] *OR* news[Title/Abstract] *OR* online [Title/Abstract] *OR* outbreak[Title/Abstract] *OR* reporting system[Title/Abstract] *OR* social [Title/Abstract] *OR* syndromic surveillance[Title/Abstract] *OR* warning system[Title/Abstract] *OR* web-based[Title/Abstract]).

The search was a combination of all the key terms generated by using the Boolean functions "*AND/OR*". These key terms were used such that Pubmed was to find them only in article "Title" or "Abstract" from the time period 1990 – 2011 (i.e. till April 7, 2011 when the search was conducted) and the search produced 9161 articles.

Due to the large number of articles found, the first stage of inclusion/ exclusion was to access the articles by reading only their titles. Titles that appeared to describe something related to surveillance were retained and titles which appeared to describe some other domain were excluded. For those titles which were difficult to be classified, their abstract were accessed, so as not to eliminate a relevant articles nor include an irrelevant article. An example of such titles that were difficult to interpret based on reading only the title was for example "The EUROMEDIES EDI prototype system".[19] By analysing these 9161 articles based on reading only the titles (and for some few cases, the abstracts), 320 articles were retained and 8841 articles were excluded.

2.4.2 Scopus

Scopus was the second database to be used and the search strategy for Scopus was different from that of Pubmed, first, because they have different subject thesauri or subject terminology and secondly, the maximum number of article's title or abstract that can be accessed using Scopus is 2000. The keywords in Scopus were combined as follows:

Surveillance *AND* infectious disease *OR* early detection *OR* early warning *OR* electronic *OR* epidemic intelligence *OR* event-based *OR* internet-based *OR* media

OR news *OR* online *OR* outbreak *OR* reporting system *OR* social *OR* syndromic surveillance OR warning system *OR* web-based

Unlike Pubmed where the articles were searched using title and abstract, Scopus was searched using, “Title, Abstracts and Keywords”. In order to limit the number of articles retrieved such that they would not exceed the maximum number that can be accessed by Scopus which was (or is) 2000, the search was performed in batches of one year interval as shown in Table 5.

Table 5 One year interval batch search using Scopus

Batch	Year	Articles	Date
1	1990	271	27.04.2011
2	1991	330	27.04.2011
3	1992	412	27.04.2011
4	1993	525	27.04.2011
5	1994	610	27.04.2011
6	1995	736	27.04.2011
7	1996	784	28.04.2011
8	1997	928	28.04.2011
9	1998	930	28.04.2011
10	1999	1126	28.04.2011
11	2000	1270	29.04.2011
12	2001	1275	29.04.2011
13	2002	1429	29.04.2011
14	2003	1700	30.04.2011
15	2004	2072	30.04.2011
16	2005	2391	01.05.2011
18	2006	2755	02.05.2011
19	2007	2841	03.05.2011
19	2008	3073	04.05.2011
20	2009	3123	05.05.2011
21	2010	3466	06.05.2011
22	2011*	777	07.05.2011

2011* was from January till current date which was May 7.

For those batches which exceeded 2000 articles, the articles were accessed by first listing them from newest down to oldest by publication date, and then inverted the order by listing them from oldest to newest. By so doing batches greater than 2000, but less than 4000 could be accessed at the 2000 limit. The total number of articles accessed using Scopus from 1990 - 2011 were 32,824. Using the first stage of inclusion/exclusion which was reading only the titles, 298 articles were retained and 32,526 articles were eliminated.

2.4.3 Scirus

The final database used was Scirus. The search was conducted on 08.05.2011 and the keywords were combined in the following manner:

Surveillance *AND* infectious disease *OR* early detection *OR* early warning *OR* electronic *OR* epidemic intelligence *OR* event-based *OR* internet-based *OR* media *OR* news *OR* online *OR* outbreak *OR* reporting system *OR* social *OR* syndromic surveillance *OR* warning system.

The number of articles highlighted by this database was 160,637,507. By default the articles are listed by relevance. Only the titles of the first 6000 articles were accessed and 79 articles were retained.

The 230 articles from Pubmed, 298 from Scopus, 79 from Scirus and 130 from the preliminary search were imported to Zotero which is an easy-to-use online reference manager and research tool that integrates tightly with online resources and helps gather, organize, and analyze sources (citations, full texts, web pages, images, and other objects).[20] The articles were listed in an alphabetical order and duplicates were checked and eliminated, thus there were 827 articles altogether and 243 were eliminated as duplicates and 584 articles were retained.

The second stage of inclusion/ exclusion was to read those 584 articles by their abstracts and to classify them in one of the categories: *"Background"* or *"Systems"*. Background articles were identified as those articles which did not directly describe an event-based surveillance system, but could be used in developing background on event-based surveillance in particular or surveillance systems in general. Hence, the reason for which they were identified as, *"Background"*. Articles in the *"System"* were those articles which described at least one system. *"System"* was distinguished between *"Monitoring System"* and "*Surveillance System*" based on the definition as shown in Table 1 (page 2).

The articles classified under *"Surveillance Systems"* were further sub-divided into *"Indicator-based*" and *"Event-based Systems"*. These sub categories were distinguished from each other based on the definitions shown in Table 2 (page 4).

2.5 Evaluation criteria used to describe results

We used a specified set of evaluation criteria to describe all of the resulting systems uncovered in this review.

The criteria we used were adopted from the work of Vrbova et al., [6] which provides several criteria for evaluating event-based surveillance systems. (Please see section 1.5 above). We also added new fields such as ‘System category’, ‘Country’, ‘I

Languages’ and system’s ‘Home page’ because we considered them important aspects of event-based surveillance systems.

Fields such as ‘Jurisdiction’, ‘Disease type’ and ‘Most avid users’ were modified to fit with the objectives of our review.

The final 15 criteria that we chose are presented under the label “Field Name” in Table 20 (please see Appendix 1). These criteria served as the basis for our extraction of relevant data on event-based systems.

3. Results

Based on the definitions of indicator-based, event-based and monitoring systems as well as our inclusion/ exclusion criteria, by reading only the abstracts of articles, 123 articles were identified as *"Background"*, 6 articles were identified as describing *"Monitoring Systems"*, 166 articles were identified as describing *"Indicator-based System"* and 44 articles were identified as describing an *"Event-based Surveillance"* system. Within those 44 articles, 18 event-based surveillance systems were identified based on reading only the abstracts. After reviewing the full texts of those 44 articles, 5 of the 18 systems that were previously classified as *"Event-based Surveillance"*, didn't have sufficient information to be extracted (Table 20) and hence were eliminated, resulting in the elimination of 12 articles (Appendix 3; Table 22) describing these 5 systems. We ended up with 32 articles (Appendix 2; Table 21) describing 13 *"Event-based Surveillance"*. Table 6 shows the list of these 13 event-based systems identified.

Table 6 List of Event-based systems identified

Nr.	System	System category	Country	Founded
1	Argus[21,22]	Moderated	USA	2004
2	BioCaster[23]	Automatic	Japan	2006
3	EpiSPIDER[24,25]	Automatic	USA	2006
4	EWRS[26]	Moderated	EU	1998
5	GOARN[27]	Moderated	USA	2000
6	GODS[N28]	Automatic	USA	
7	GPHIN[29,30]	Moderated	Canada	1997
8	HealthMap[31–38]	Automatic	USA	2006
9	InSTEDD[39]	Moderated	USA	2006
10	MedISys and PULS[40–42]	Automatic	EU	2004
11	MiTAP[43]	Automatic	USA	2001
12	ProMED[44–51]	Moderated	USA	1994
13	Proteus-BIO[52]	Automatic	USA	

3.1 General Results

In the following section we summarize all of the identified event-based surveillance systems and present general results from our evaluation of each of our 15 criteria.

3.1.1 System category

Event-based surveillance systems can be classified into one of these 3 categories; news aggregators, automatic and moderated systems.[8]

News aggregators collect articles from several sources, usually filtered by language or country. Users gain easy access to many sources through a common portal, but still need to examine each individual article.

Automatic systems go beyond the mere gathering task by adding a series of analysis steps. Automatic systems differ in their levels of analysis, in the range of information sources, their language coverage, the speed of delivering information and visualisation methods.

Moderated systems rely on a group of (human) analysts to scan available news sources. The analysts take into account information from individual web sites, aggregator sites, automatic systems, and other sources such as reports from medical practitioners and health authorities.

In our review, we found only automatic and moderated systems. There are 7 automatic systems (54%) and 6 moderated systems (46%). The automatic systems include; BioCaster, EpiSPIDER, GODS[N], HealthMap, MedISys-PULS, MiTAP and Proteus-BIO. The moderated systems include; Argus, EWRS, GOARN, GPHIN, InSTEDD and ProMED.

These system categories have fundamental differences in their approaches. For example, non-moderated systems are able to search the web and display new articles without time delay in an unbiased manner than moderated systems. On the other hand, moderated systems might show fewer irrelevant news items (fewer false positives) than the non-moderated systems. However, moderator bias represents a risk (false negatives); users might have a different focus than the moderators.[8]

3.1.2 Coordinating organization

Of all the event-based systems, we identified three types of coordinating bodies; a university-based coordinated system, an NGO-coordinated system and a governmental agency coordinated system. 5 systems (38%) are university-based systems which include; Argus, BioCaster, GODS^N, HealthMap and Proteus-BIO. Also 5 systems (38%) are coordinated by an NGO and include; EWRS, GOARN, MedISys, MiTAP and ProMED. Only 1 system (GPHIN) is coordinated by a governmental agency. We were not able to identify the coordinating body of 2 systems (EpiSPIDER and InSTEDD).

Surveillance systems based in a university or NGO are not subject to government constraints on information flow, but when moderated by health professionals maintain high credibility.[45]

3.1.3 Purpose

The purposes of these identified systems can be classified into three different categories; 1) those that aim to enhance early detection, 2) those that aim to enhance communication or collaboration and 3) those that aim to supplement other existing system(s). 10 of the systems (77%) aim to enhance early detection and these include; Argus, BioCaster, GOARN, GODS^N, GPHIN, HealthMap, InSTEDD, MedISys, MiTAP and Proteus-BIO. 2 of the systems (15%) aim to enhance communication or collaboration and these include; EWRS and ProMED. 1 system (8%) aim to supplement other existing system (ProMED) and this includes EpiSPIDER.

3.1.4 Jurisdiction

All the systems operate at the international level i.e covers 2 or more countries, but their jurisdictions could be classified in one of these 3 categories; 1) those that monitor worldwide, 2) those that give preference only to a confined region and 3) those that monitor any other region apart from the region where the system is based. 9 of the systems (69%) monitor worldwide and these include; EpiSPIDER, GOARN, GODS^N, GPHIN, HealthMap, InSTEDD, MiTAP, ProMED and Proteus-BIO. 3 systems (23%) pay more attention to a confined region which include BioCaster (mostly Asia-Pacific countries), EWRS (only in the EU and the EEA area) and MedISys (other regions, but more particularly Europe). Only Argus which is a US-based system monitors elsewhere, but excludes monitoring in the US. The reason for this exception was not stated.

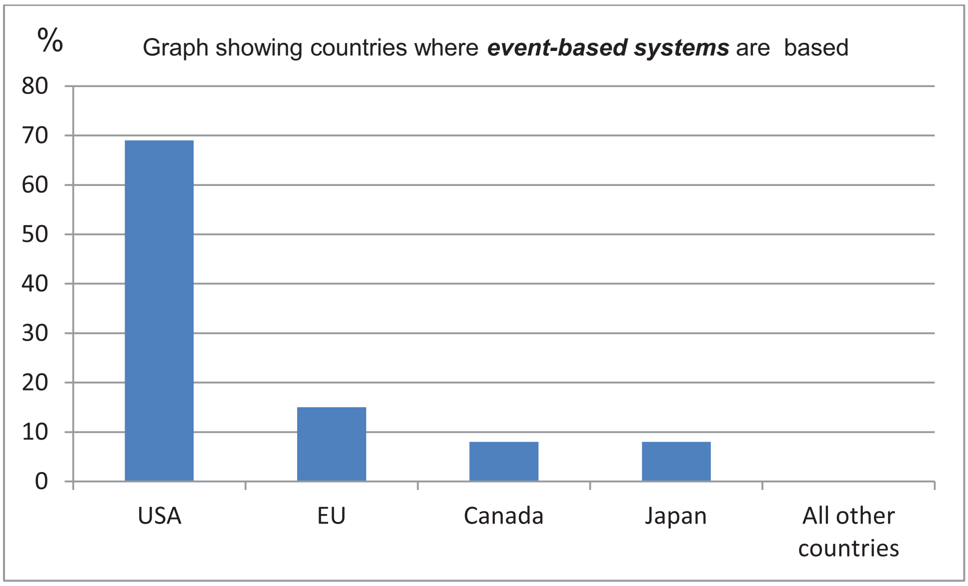

Figure 4 Percentage representation of countries with event-based systems

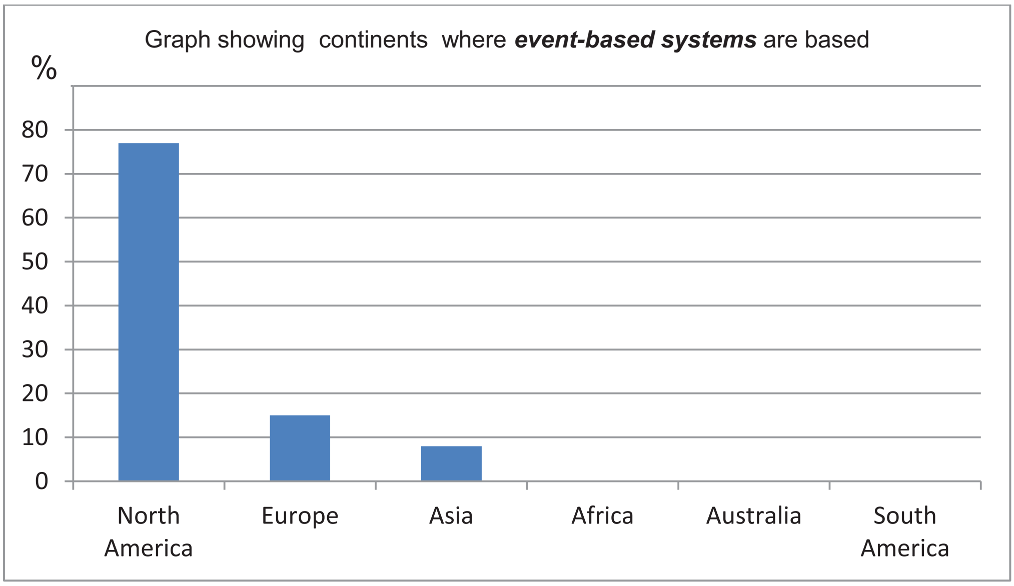

Figure 5 Percentage representation of continents with event-based systems

3.1.5 Languages

There are two distinguished language categories in which the event-based systems either collected or disseminated their data; first category is "only in 1 language" and the second category is "2 or more languages". 5 of the systems (38%) used "only 1 language" and these include; EpiSPIDER, EWRS, GODS[N], InSTEDD and Proteus-BIO. 8 of the systems (62%) use "2 or more languages" and include Argus (34 languages), BioCaster (8 languages), GPHIN (8 languages), HealthMap (5 languages), MedISys (43 languages), MiTAP (8 languages) and ProMED (7 languages).

3.1.6 Disease type

All the event-based systems focused in the outbreak of multiple infectious diseases. When a system collects three or more disease entities, we classified it as multiple infectious diseases. Some systems collect over 100 disease entities which include for example Argus (over 130 disease entities), BioCaster (approx. 102 disease entities) and HealthMap (approx. 170 disease entities).

3.1.7 Public Access

The public access can be grouped into 5 categories; 1) freely available, 2) free subscription, 3) paid subscription, 4) restricted and 5) those that provided access at multiple levels. 4 systems are freely available to the public and these include; HealthMap, EpiSpider, $GODS^N$, and Proteus-BIO. 3 systems offer free subscription to anyone and these include; ProMED, BioCaster and MiTAP. GPHIN is the only system which offers paid subscription.subscribers mostly indlude; Government of Canada Organizations, Non-Canadian government organizations, International Organizations, Non-Profit Organizations, Universities and Research Institutions. Four systems provide restricted access only to certain officials (mostly public health officials) and these include EWRS, Argus, GOARN and InSTEDD. Only MedISys provides access at multiple levels where are; 1) free public access, 2) restricted access for public health professionals outside the EC and 3) full access inside the EC.

3.1.8 Data Acquisition

The manner in which these event-based systems acquire their data can be classified in three categories; 1) those that collect from either internet, RSS feeds or electronic mailing list(s), 2) those that collect both from formal and informal sources and 3) those that collect from subscribers. 10 systems (77%) collect from either internet, RSS feeds or electronic mailing list(s) for which include Argus, BioCaster, EpiSPIDER, $GODS^N$, GPHIN, HealthMap, InSTEDD, MedISys, MiTAP and Proteus-BIO. 2 systems (15%) collect from both formal and informal sources and these include; EWRS and GOARN. ProMED is the only system which obtains its first-hand information from its subscribers.

3.1.9 Data Processing

For the 13 event-based systems, we identified 2 different ways how data are being processed. These include automatic and moderated processing (see System category; page 16). For automatic systems, data processing is done entirely by machine and for moderated systems, data is either done entirely by human analysts or processing might begin with machine and later passed on to human analysts for moderation.

3.1.10 Dissemination of Data

There are three identified categories, in which the data of event-based systems are being disseminated, 1) through a secured or restricted portal, 2) laid on a geographic map and 3) to a website or newsgroup. 6 systems (46%) were found to disseminate through a secured or restricted portal and these include; Argus, EWRS, GOARN, GODS[N], GPHIN and InSTEDD. 3 systems (23%) disseminate on a geographic map which include; BioCaster, EpiSPIDER and HealthMap. 4 systems (31%) disseminate through a website or newsgroup which include; MedISys, MiTAP, ProMED and Proteus-BIO.

3.2 Results on individual systems

In the following section each identified event-based surveillance system is presented in greater detail. The full criteria are shown in respective tables, and a closer analysis of the following criteria is expanded upon: Data Acquisition, Data Processing, and Data Dissemination. I chose to expand upon those 3 criteria because these are the 3 basic units that make up a surveillance system. For a system to be identified as a surveillance system, then it should be able to acquire data, processes the data and disseminates the data in a timely manner (for rapid intervention if necessary).

3.2.1 Argus

Table 7 Overview of Argus

No	Field Name	Field Description
1	System name	Argus
2	System category	Moderated system
3	Country	USA
4	Year started	2004
5	Coordinating organization	Georgetown University Medical Center (Washington DC, United States).
6	Purpose	To create and implement a global biological event detection and tracking capability that provides early warning alerts.
7	Jurisdiction	International (approx. 175 countries excluding the US).

No	Field Name	Field Description
8	Languages	34
9	Disease type	Multiple infectious diseases (over 130 disease entities).
10	Public access	Restricted
11	Data processing	Argus utilizes a proprietary state of the art online media processing software with a taxonomy of nearly 200 indicators coupled to a heuristic staging scale for the identification of biological events. Analysts then evaluate the report once an event is identified for possible posting as a warning, watch, or advisory.
12	Dissemination of data	The reports are posted on a secure Internet portal for the diverse set of Argus users to view.
13	Most avid users	WHO, Department of Homeland SecurityHealt and Human Services, CDC, USDA
14	System evaluation	Yes[21]
15	Homepage	http://biodefense.georgetown.edu/

The activation of Project Argus in late 2004 was supported by the Department of Homeland Security (DHS) and the Intelligence Technology Innovation Center (ITIC) due in part to the emergence of SARS in southern China in 2002 and 2003 as well as the Highly Pathogenic Avian Influenza–H5N1 pandemic concern. The Argus system, is a web-based global biosurveillance system hosted at the Georgetown University Medical Center (Washington DC, United States) designed to report and track the evolution of biological events threatening human, plant and animal health globally, excluding the United States.[53]

Data Acquisition

Argus covers 34 languages with a team consisting of multi-lingual analysts that utilize proprietary state of the art online media processing software designed in collaboration with the MITRE Corporation which monitors several thousand internet sources six times daily and collects in an automated process, local, native-language internet media reports, including blogs and official sources e.g. World Health Organization (WHO) and World Organisation for Animal Health (OIE). They use Boolean keyword searching and Bayesian model tools to select reports from a dynamic database of media reports.[22]

Data Processing

A taxonomy of nearly 200 indicators coupled to a heuristic staging model called the Wilson-Collmann Scale is used for the identification of biological events.[21]

The project analysts evaluate and write event reports, which are based on relevant media reports, and a stage is assigned to the report based on observed event progression according to a previously described heuristic model ranging from preparatory to degree of disease spread to degrees of social disruption to recovery and interprets their relevance according to a specific set of concepts and keywords relevant to infectious disease surveillance.[22]

Dissemination of data

Using a disease event warning system modelled after NOAA's National Weather Service, the reports are posted on a secure internet portal as a warning, watch, or advisory for the diverse set of Argus users to view. These warnings, watches, and advisories are posted in accordance with guidelines agreed upon by their research partners in the federal government. On average, they maintained 15 Advisories, 5 Watches, and 2 Warnings active on their Watch board at any given time. Argus currently manages between 2,200 to 3,300 active, socially disruptive biological event case files with update report threading for approximately 175 countries (except the United States) and over 130 disease entities resulting in the production of, on average, 200 reports per day.[53]

3.2.2 BioCaster

Table 8 Overview of BioCaster

No	Field Name	Field Description
1	System name	BioCaster
2	System category	Automatic system
3	Country	Japan
4	Year started	2006
5	Coordinating organization	National Institute of Informatics, Tokyo
6	Purpose	An early warning system developed to enhance early detection of epidemic and environmental diseases (human, animal and plant) by experts.
7	Jurisdiction	International (Asia-Pacific countries)
8	Languages	8 (Chinese, English, Japanese, French, Korean, Spanish, Thai and Vietnamese).
9	Disease type	Multiple infectious diseases (approx. 102 infectious diseases)
10	Public access	Free subscription

No	Field Name	Field Description
11	Data processing	BioCaster continuously scans hundreds of RSS newsfeeds from local and national news providers, aggregates these online news reports, processes them automatically using human language technology and tries to spot unusual trends and classifies them for topical relevance.
12	Dissemination of data	• More precisely specified warning signals are notified to registered users via email alerts. • Events are also plotted onto Google map using geocoded information.
13	Most avid users	National Institute of Infectious Diseases, National Institute of Genetics, Okayama University, Vietnamese National University at Ho Chi Minh City and Kasetsart University.
14	System evaluation	Yes[54,55]
15	Homepage	http://born.nii.ac.jp/

BioCaster is an ontology-based text mining non-governmental public health surveillance system established in 2006 for detecting and tracking the distribution of infectious disease outbreaks from linguistic signals on the Web. This system is characterized by its open ontology-centered approach and a priority for Asia-Pacific languages and health hazards. The system consists of four main stages: topic classification, named entity recognition (NER), disease/location detection and event recognition.[23]

Data Acquisition

The BioCaster system has two major components: a web/database server and a backend cluster computer equipped with a variety of text mining algorithms which continuously scans and analyzes documents reported from over 1700 RSS newsfeeds from local and national news providers[56] for which the sourced documents are then cleansed and put into the cluster queue. Automatic classification of the reports for topical relevance using a naïve Bayes algorithm then acts as the gate-keeper for further levels of processing.[23]

Data Processing

At the core of BioCaster is the BioCaster ontology (BCO), developed by a multi-disciplinary team of experts. The BCO is organized around an application taxonomy with root terms representing key domain concepts. The BCO encoded as an openly available Web Ontology Language (OWL) file gives access to term definitions, synonyms and translations in eight languages as well as mappings to external ontologies.

Named entity recognition (NER) is then performed for 18 term types based on the BCO for relevant documents.[57]

Analysis then focuses on detecting domain-specific signals, such as cases of drug resistance, malformed blood products, international travel, zoonosis or newly emerging strains.[23]

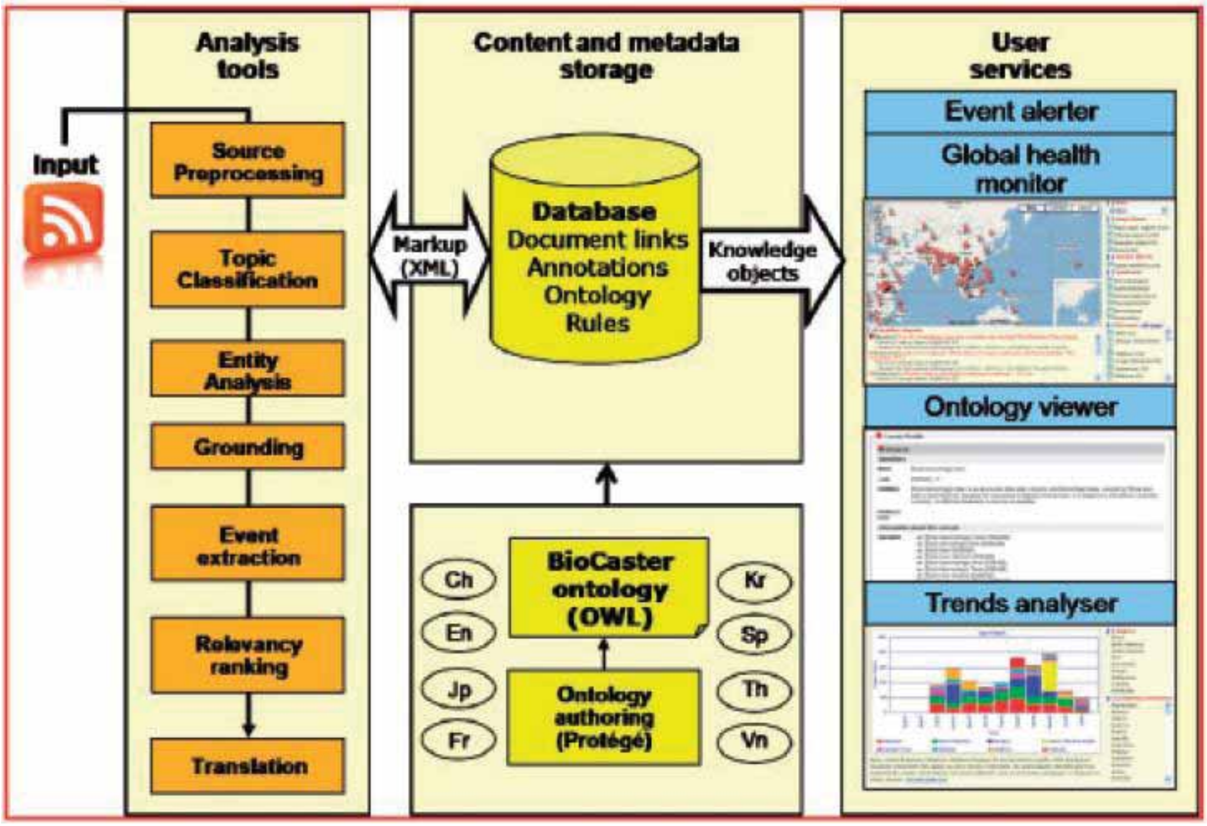

Figure 6 Overview of BioCaster system showing the stages of text-to-knowledge conversion followed by user service provision.[23]

Data Dissemination

The text mining system has a detailed knowledge about the important concepts such as diseases, pathogens, symptoms, people, places, drugs etc. this allows BioCaster to semantically index relevant parts of news articles and classifies events for topical relevance and disease-location pairs are plotted onto a public portal called the Global Health Monitor, using geocoded information. This enables users to have quicker and highly precise access to information and gain a geographically contextualized view of an outbreak anywhere in the world which can be filtered by pathogen, syndrome or text type. Users can drill down to source evidence by clicking on map points which display associated headlines for the event along with links to scientific databases, such as PubMed, HighWire and Google Scholar.

Higher order event analysis is used to detect more precisely specified warning signals that can then be notified to registered users via email alerts on topics of interest. BioCaster includes information on 102 infectious diseases as well as geo-locations for two administrative levels.[23]

3.2.3 Semantic Processing and Integration of Distributed Electronic Resources for Epidemiology (EpiSPIDER)

Table 9 Overview of EpiSPIDER

No	Field Name	Field Description
1	System name	Semantic Processing and Integration of Distributed Electronic Resources for Epidemiology (EpiSPIDER).
2	System category	Automatic system
3	Country	USA
4	Year started	2006
5	Coordinating organization	
6	Purpose	The EpiSPIDER project was designed to serve as a visualization supplement to the ProMED-mail reports.
7	Jurisdiction	International (approx. 50 countries).
8	Languages	English
9	Disease type	The number covered by ProMED
10	Public access	Freely available
11	Data processing	EpiSPIDER obtains information from sources like ProMED mailing list, RSS feeds from health news like infectious disease sections of the WHO, European Surveillance Network, GDACS web link and news syndication sites like Reuters. EpiSPIDER extracts location information from these sources using natural language processing and geocode them using the Yahoo and Google geocoding services.
12	Dissemination of data	It publishes content on Google Map and makes available filtered and unfiltered RSS feeds for downloading to agencies and organizations.
13	Most avid users	Mostly persons in North America, Europe, Australia, and Asia, CDC, USDA, US Department of Homeland Security, US Directorate for National Intelligence, UK Health Protection Agency, several universities and health research organizations.
14	System evaluation	Not evaluated
15	Homepage	http://www.epispider.org/

The EpiSPIDER project was designed in January 2006 to serve as a visualization supplement to the ProMED-mail reports.[25] It is an integrative web-based information processing system that uses electronic resources to create an information environment for enhancing the surveillance of emerging infectious disease threats to global health.[24] It uses both the Google Maps and Yahoo Maps APIs. EpiSPIDER generates country-level maps for all countries covered by ProMED reporting, as well as state-level maps for the US, Canada and a few selected countries.[25]

Data Acquisition

EpiSPIDER uses the mailing list of ProMED as a source of emerging infectious disease reports. Other sources come from Really Simple Syndication (RSS) feeds health news like infectious disease sections of the WHO, the European Surveillance Network, the Global Disaster Alert Coordinating System (GDACS). EpiSPIDER also obtains RSS feeds containing health information from news syndication sites like Reuters.

For ProMED reports, the following fields are extracted: date of publication; list of locations (country, province, or city) mentioned in the report; and topic. Each news report that has location information can be linked to relevant demographic- and health-specific information (e.g., population, per capita gross domestic product, public health expenditure, and physicians/1,000 population).[25]

Data Processing

EpiSPIDER extracts location information from these sources using natural language processing and geocode them using the Yahoo and Google geocoding services.[24] Additonally, EpiSPIDER automatically converted the topic and location information of the reports into RSS and other formats such as keyhole markup language (KML) and JavaScript object notation (JSON) which is a human-readable format for representing simple data structures.[25]

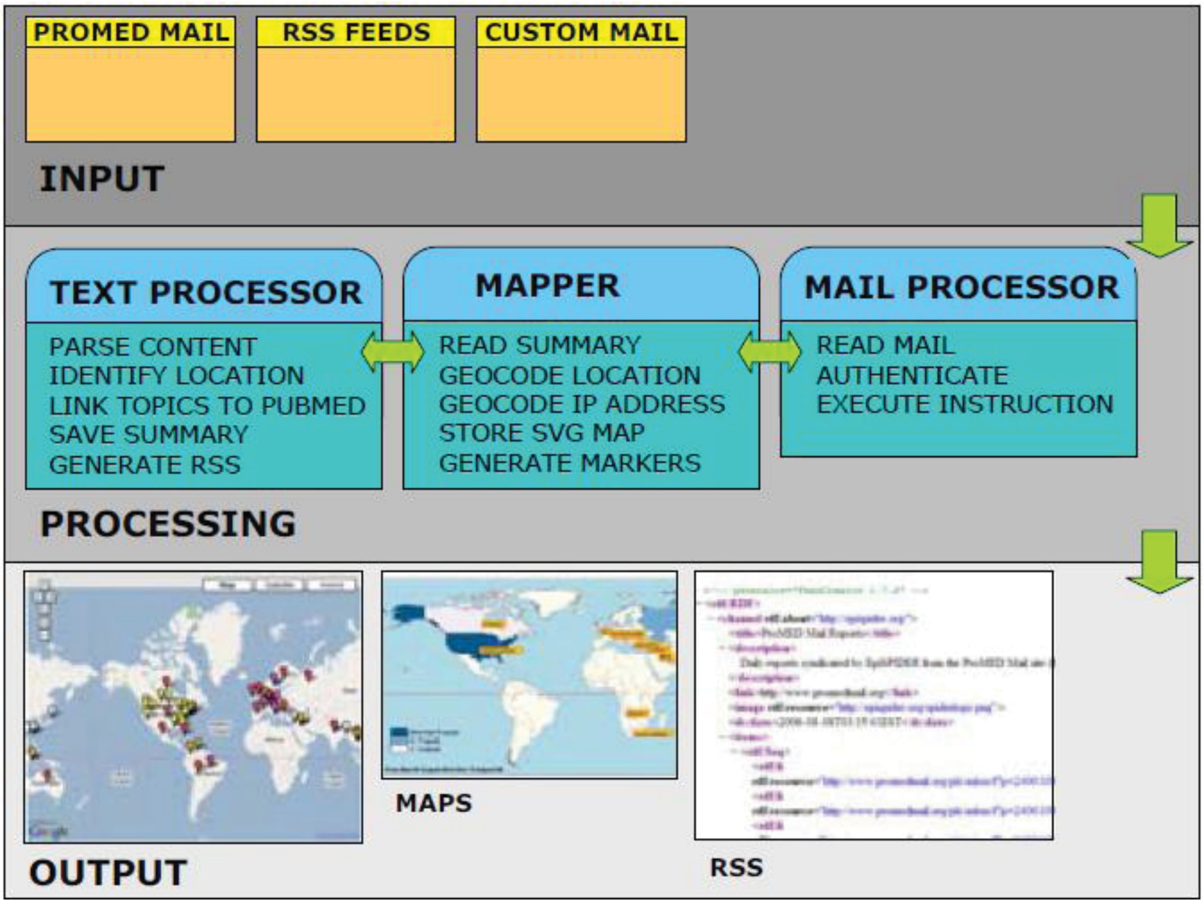

Figure 7 Inputs, processing and outputs of EpiSPIDER. EpiSPIDER processes custom mail instructions through its native mail processor.[24]

Data Dissemination

Publishing content using those formats like RSS, KML and JSON enables the semantic linking of ProMED-mail content to country information and facilitates EpiSPIDER's redistribution of structured data to services that can consume them.

As a data provider, it makes available these filtered and unfiltered RSS feeds for downloading by agencies and organizations.

EpiSPIDER is used mostly by persons in North America, Europe, Australia, and Asia, and it receives 50–90 visits/hour, originating from 150–200 sites and representing 30–50 countries worldwide. EpiSPIDER has recorded daily visits from the US Department of Agriculture, US Department of Homeland Security, US Directorate for National Intelligence, US CDC, UK Health Protection Agency, and several universities and health research organizations.[25]

3.2.4 Early Warning and Response System (EWRS)

Table 10 Overview of EWRS

No	Field Name	Field Description
1	System name	Early Warning and Response System (EWRS)
2	System category	Moderated system
3	Country	European Union
4	Year started	1998
5	Coordinating organization	ECDC (since 2007)
6	Purpose	To promote cooperation and coordination between the Member States, with the assistance of the European Commission, with a view to improving the prevention and control of communicable diseases related events.
7	Jurisdiction	The EU Member States, Bulgaria, Romania and the European Economic Area countries (Iceland, Liechtenstein and Norway).
8	Languages	English
9	Disease type	Multiple infectious diseases
10	Public access	Restricted
11	Data processing	In addition to formal sources, the EWRS is linked to the MedISys which is an informal source. Access to the EWRS is secured and is limited to the formally appointed contact points. Following notification, the contact point receives a login and a password from the Commission to access the system, and full authorisation to read and write messages.
12	Dissemination of data	When a message is posted on the system, it is automatically circulated to all EWRS contact points and the entire EWRS network of how the situation is progressing and of the measures planned or undertaken at national level to respond to the specific event.
13	Most avid users	The national health authorities, the national public health agencies, the Ministries for Health in Member States, and the European Commission and its agencies, in particular the ECDC.
14	System evaluation	Yes[58]
15	Homepage	https://ewrs.ecdc.europa.eu/

A network for epidemiological surveillance and control of communicable diseases in the European Union (EU) was set up in 1998 by the European Parliament and of the Council. One pillar of Decision 2119/98/EC was to setup an early warning and response system (EWRS) between Member States. The main objective of the network was to establish permanent communication between EU Member States' public health authorities, which are responsible for determining the measures required to control communicable disease-related events. A web based informatics tool has been developed in order to allow information to be shared between the relevant public health authorities.

Data Acquisition

In addition to obtaining events from formal sources, the EWRS is linked to the Medical Intelligence System (MedISys) which is an informal source. MedISys is a piece of software that browses the web every 20 minutes in order to find articles, documents and latest news about health matters. Approximately 1200 websites are visited with about 350 keywords are currently used. Access to the EWRS is secured and is limited to the formally appointed contact points.

Data Processing

An informatics tool has provided the platform for communicating information and a web-based system currently links the EWRS contact points. Access to the system is secured and is limited to the formally appointed contact points. Following notification from Member States, the contact point receives a login and a password from the Commission to access the system, and full authorisation to read and write messages.

Data Dissemination

When a message is posted on the system, it is automatically circulated to all EWRS contact points, and the network (Commission, Member States, acceding and the EEA countries, and ECDC) is informed at the same time of how the situation is progressing and of the measures planned or undertaken at national level to respond to the specific event. Access to the EWRS has been granted to Ministries of Health, national surveillance institutes, specific EU supported projects, ECDC and WHO.

A short message service (SMS) messaging function has been activated in order to transmit to the European Commission Officer on duty real time notification that a message has been posted on the system. Graphs, statistics and world maps allow a quick identification of threats and localisation of the events.

From the time point between 1998 and December 2005, a total of 583 messages were circulated through the EWRS. These messages notified a total of 396 events.[26]

3.2.5 Global Outbreak and Alert Response Network (GOARN)

Table 11 Overview of GOARN

No	Field Name	Field Description
1	System name	Global Outbreak and Alert Response Network (GOARN)
2	System category	Moderated system
3	Country	USA
4	Year started	2000

No	Field Name	Field Description
5	Coordinating organization	WHO
6	Purpose	To establish an infrastructure for responding to the heightened need for early awareness of outbreaks and preparedness to respond.
7	Jurisdiction	International (worldwide)
8	Languages	English & French
9	Disease type	Multiple infectious diseases
10	Public access	Restricted
11	Data processing	The procedure for outbreak alert and response has four phases: systematic detection, outbreak verification, real time alerts, and rapid response.
12	Dissemination of data	An electronic Outbreak Verification List is distributed on a weekly basis to over 800 individuals within the 110 networks and including staff in the national quarantine offices of WHO member countries.
13	Most avid users	Government and university centres such as CDC, the UK Public Health Laboratory Service, the French Institute Pasteur outbreaks, UN agencies (notably United Nations High Commissioner for Refugees and Unicef), ministries of health, academic institutes, WHO regional and country offices, and government networks of overseas military laboratories, such as the US Department of Defence Global Emerging Infections System (GEIS), GPHIN and over 250 laboratories and institutions formally designated as WHO collaborating centres.
14	System evaluation	Yes[59]
15	Homepage	http://www.who.int/csr/outbreaknetwork/en/

In April 2000, WHO formalised GOARN which is an infrastructure for responding to the heightened need for early awareness of outbreaks and preparedness to respond. GOARN's network unites 110 existing networks and supported by several new mechanisms and a computer-driven tool for real time gathering of disease intelligence which together possess much of the data, expertise, and skills needed to keep the international community constantly alert and ready to respond.

Data Acquisition

GOARN collects part of its information from informal sources such as GPHIN which continuously and systematically crawling web sites, news wires, local online newspapers, public health email services, and electronic discussion groups, including ProMED-mail. In this way, WHO is able to scan the world for informal news that gives cause for suspecting an unusual disease event. Other informal reports also come in from non-governmental organisations, such as the Red Cross and Crescent societies, Médecins Sans Frontières, and religious organisations, such as the Catholic and Protestant mission networks

Data Processing

Raw intelligence collected from all formal and informal sources is converted into meaningful intelligence by the WHO Outbreak Verification Team, which meets daily to review incoming reports and rumours, assess their epidemiological significance, and decide on the actions needed. Four main criteria are used to determine whether an event is of potential international public health importance: serious health impact or unexpectedly high rates of illness and death, potential for spread beyond national borders, potential for interference with international travel or trade, and strength of national capacity to contain the outbreak.[60] The team also routinely considers whether reports of an unusual disease event might be associated with deliberate or accidental release of a biological agent.

Data Dissemination

A detailed standardised report on suspected and verified outbreaks is distributed electronically at the end of the day to a limited number of WHO staff at headquarters and in regional offices. The WHO team distributes, on a weekly basis, an electronic Outbreak Verification List to over 800 individuals within the 110 networks and including staff in the national quarantine offices of WHO member countries. The list provides restricted access to news of both initial, unconfirmed, and potentially sensitive reports of outbreaks and the status of confirmed outbreaks undergoing investigation. Once an outbreak has been verified, WHO posts situation reports on its web site and publishes them in the Weekly Epidemiological Record, which is available electronically, in English and French, and distributed in printed form to a large number of recipients in the developing world. WHO's six regional offices and 141 country offices in less-developed countries provide an important geographical resource for coordinating containment operations within countries and their neighbours. All offices are staffed with medical experts and often with epidemiologists, and all have the essential logistic equipment, including vehicles and local communications, needed for prompt on-the-scene investigation of a suspected outbreak. Investigative teams from WHO Headquarters in Geneva are prepared to arrive at an outbreak site within 24 h.

By electronically linking together existing networks, WHO is thus able to magnify its limited resources considerably.[27]

3.2.6 Global News Driven Disease Outbreak and Surveillance (GODS^{N})

Table 12 Overview of GODS^{N}

No	Field Name	Field Description
1	System name	Global News Driven Disease Outbreak and Surveillance (GODS^{N}) read as n-GODs
2	System category	Automatic system
3	Country	USA
4	Year started	
5	Coordinating organization	Department of Biomedical Informatics, Columbia University, New York, NY
6	Purpose	To monitor global news for disease outbreaks and surveillance and process real-time news feeds to extract specific diseases (or healthcare) events and their location of occurrence.
7	Jurisdiction	International (at least in all the continents)
8	Languages	English
9	Disease type	Multiple infectious diseases
10	Public access	Freely available
11	Data processing	GODS^{N} uses publicly available web services or news aggregators that use RSS to collect real-time news feeds from thousands of news sources around the world. A natural language processing system processes the feeds to extract relevant concepts such as disease names and the reference to a geographic location.
12	Dissemination of data	The extracted information is plotted on a GIS. The Web interface of the system allows users to query for diseases and view their temporal and spatial evolution.
13	Most avid users	
14	System evaluation	Not evaluated
15	Homepage	

GODS^{N} is currently an early-prototype under active development, developed by the Department of Biomedical Informatics, Columbia University, New York, NY. GODS^{N} monitors global news for disease outbreaks and surveillance and provides an effective approach to visualize the spatial and temporal trends of infectious disease outbreaks or disease specific developments.

Data Acquisition

GODS^{N} uses publicly available web services or news aggregators such as Google^{TM} Maps and Google^{TM} news that use RSS to collect real-time news feeds from thousands of news sources around the world.

Data Processing

Using open application programming interfaces (APIs), the obtained information are passed through a filter to extract health-related stories. A natural language processing system (MetaMap) processes the feeds to extract relevant concepts such as disease names and the reference to a geographic location.[28]

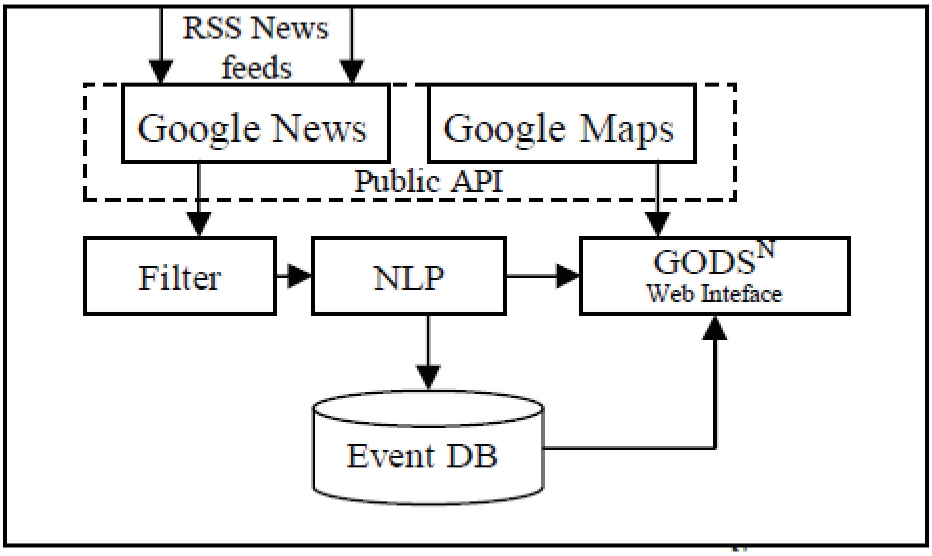

Figure 8 System Architecture of the GODS^N system[28]

Data Dissemination

The extracted information is plotted on a geographic information system (GIS). The Web interface of the system allows users to query for diseases and view their temporal and spatial evolution.

3.2.7 Global Public Health Intelligence Network (GPHIN)

Table 13 Overview of GPHIN

No	Field Name	Field Description
1	System name	Global Public Health Intelligence Network (GPHIN)
2	System category	Moderated system
3	Country	Canada
4	Year started	1997 (prototype system); 2004 (launched at UN).
5	Coordinating organization	Public Health Agency Canada - Health Canada's Centre for Emergency Preparedness and Response (CEPR).
6	Purpose	To determine the feasibility and effectiveness of using news media sources to continuously gather information about possible disease outbreaks worldwide and to rapidly alert international bodies of such events.
7	Jurisdiction	International (about 132 countries).
8	Languages	8 (Arabic Chinese simplified & traditional, English, Farsi French, Russian Portuguese and Spanish).
9	Disease type	Multiple infectious diseases (and others activities such as contaminated food and water, bio-terrorism, exposure to chemicals, and natural disasters).
10	Public access	Paid subscription

No	Field Name	Field Description
11	Data processing	GPHIN monitors global media sources such as news wires, web sites, local and national newspapers retrieved through news aggregators in 8 languages with the aid of a machine translation. The information is filtered for relevancy by an automated process which is then complemented by human analysis.
12	Dissemination of data	Data is disseminated through the internet to users such as non-governmental agencies, organizations, as well as government authorities who conduct public health surveillance. Notifications about events that might have serious public health consequences are immediately sent by email to users in the form of an alert. Organizations such as WHO, CDC and FAO use GPHIN on a daily basis. These organizations can then quickly react to public health incidents.
13	Most avid users	WHO, international government and NGO's
14	System evaluation	Yes[27]
15	Homepage	http://www.phac-aspc.gc.ca/gphin/

GPHIN is a secure Internet-based "early warning" system that gathers preliminary reports of public health significance on a "real-time" basis, 24 hours a day, 7 days a week. The prototype GPHIN system was developed in 1997 in partnership with the World Health Organization (WHO). The objective was to determine the feasibility and effectiveness of using news media sources to continuously gather information about possible disease outbreaks worldwide and to rapidly alert international bodies of such events. These informal sources included websites, news wires, and local and national newspapers retrieved through news aggregators in English and French. After haven proven effective during the severe acute respiratory syndrome (SARS) in 2002, a new, robust, multilingual GPHIN system (GPHIN II) was developed and was launched November 17, 2004, at the United Nations.

Data Acquisition

The GPHIN software application retrieves relevant articles every 15 minutes, 24 hours a day and 7 days a week from news-feed aggregators such as Al Bawaba (www.albawaba.com) and Factiva (www.factiva.com) according to established search queries that are updated regularly. These services operate as news aggregators that provide multiple sources of information through a single access point. Factiva, for example, aggregates news information from nearly 9,000 sources in 22 languages.[30]

Data Processing

Non-English articles are machine-translated into English. The articles are then filtered and categorized into one or more of GPHIN's taxonomy categories: animal diseases; human diseases; plant diseases; biologics; natural disasters; chemical incidents; radiologic incidents; and unsafe products. Each article is assigned a relevancy score within the taxonomy or taxonomies it has been assigned to.

Articles whose relevancy is below an established threshold are automatically trashed as being not relevant according to GPHIN criteria. Articles with a relevancy score above a certain threshold are automatically published to the GPHIN database. Articles that are of an even higher relevancy are also immediately sent to GPHIN users by email as alerts. Articles whose relevancy lies in the zone between the automatic publish and the automatic trash thresholds are presented to a GPHIN Analyst for human decision making who review the articles and decided whether to publish, trash or alert the article. The analysts also review the automatically 'trashed' articles to ensure that there are no relevant articles or false negatives. The team of GPHIN analysts conduct more in-depth analysis identifying and reporting on trends and assessing the health risks to the populations around the world. The GPHIN infrastructure can be seen in Figure 9 below.

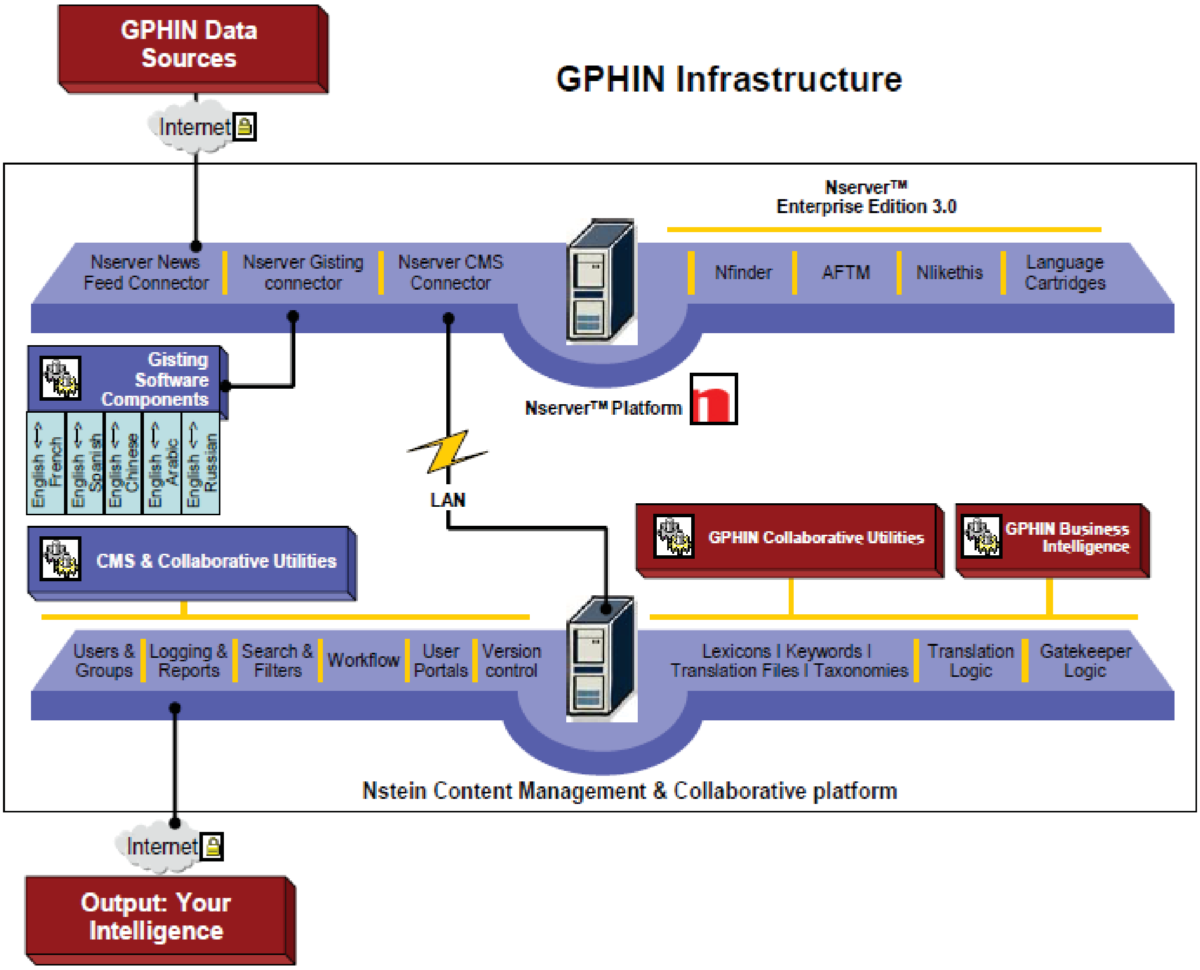

Figure 9 GPHIN infrastructure[29]

Data Dissemination

English articles are machine-translated into Arabic, Chinese (simplified and traditional), Farsi, French, Russian, Portuguese, and Spanish. Non-English articles are machine-translated into English.[25]

Subscription is restricted to organizations with an established public health mandate and varies according to such factors as organizational size and the number of users. Subscribers access GPHIN information through a restricted website. Upon accessing GPHIN, the users may review the latest list of published articles or they can further filter the list with the use of a query function to view specific articles. In addition, notifi cations about events that might have serious public health consequences immediately sent by email to users in the form of an alert.

GPHIN proved to be effective and efficient in the timely reporting of potential disease outbreaks. During the period of July 1998 to August 2001, WHO verified 578 outbreaks, of which 56% were initially picked up by GPHIN. The outbreaks reported were occurred in about 132 countries around the world demonstrating GPHIN's capacity to monitor worldwide.[27]

GPHIN processes anywhere from 2,000 to 3,000 news items per day, of which approximately 1/4 to 1/3 are discarded as irrelevant or duplicates.[30]

3.2.8 HealthMap

Table 14 Overview of HealthMap

No	Field Name	Field Description
1	System name	HealthMap
2	System category	Automatic system
3	Country	USA
4	Year started	2006
5	Coordinating organization	Harvard Medical School, Boston, MA, USA.
6	Purpose	To deliver real-time intelligence on a broad range of emerging infectious diseases to government agencies and public-health officials, as well as international travellers and local health departments.
7	Jurisdiction	International (about 160 countries)
8	Languages	5 (English, French, Chinese, Spanish & Russian. Additional languages under development include Hindi, Portuguese, and Arabic).
9	Disease type	Multiple infectious diseases (about 170 disease categories):
10	Public access	Freely available
11	Data processing	HealthMap collects and integrates outbreak data from a variety of electronic sources such as Google News, ProMED Mail and WHO validated official alerts. The system classifies alerts by location and disease through the use of automated text processing algorithms. Once classified, articles are filtered by their relevance into five categories: Breaking News, Warning, Follow-up, Background/Context and Not Disease Related.
12	Dissemination of data	Only "breaking news" articles are overlaid on an interactive geographic map.
13	Most avid users	WHO, CDC, ECDC, UN, FDA, Italy's national epidemiology agency, US Health and Human Services and the US Department and other national, state and local bodies around the world.
14	System evaluation	Yes[33,35]
15	Homepage	http://www.healthmap.org/en/

HealthMap is a freely accessible, automated web-based real-time system operating since September 2006 designed to collect and visualise outbreak data according to geography, time, and infectious disease agent. Its goal is to deliver real-time intelligence on a broad range of emerging infectious diseases for a diverse audience, from public health officials to international travellers.[38] HealthMap provides access to the greatest amount of potentially useful health information across a wide range of geography and pathogens, without overwhelming the user with excess information or obscuring important and urgent elements.[35]

Data Acquisition

The system is made up of five modules, which are the Data Acquisition Engine, the Classification Engine, the Database, the Web Backend, and the Web Frontend. The system architecture can be seen in Figure 10.

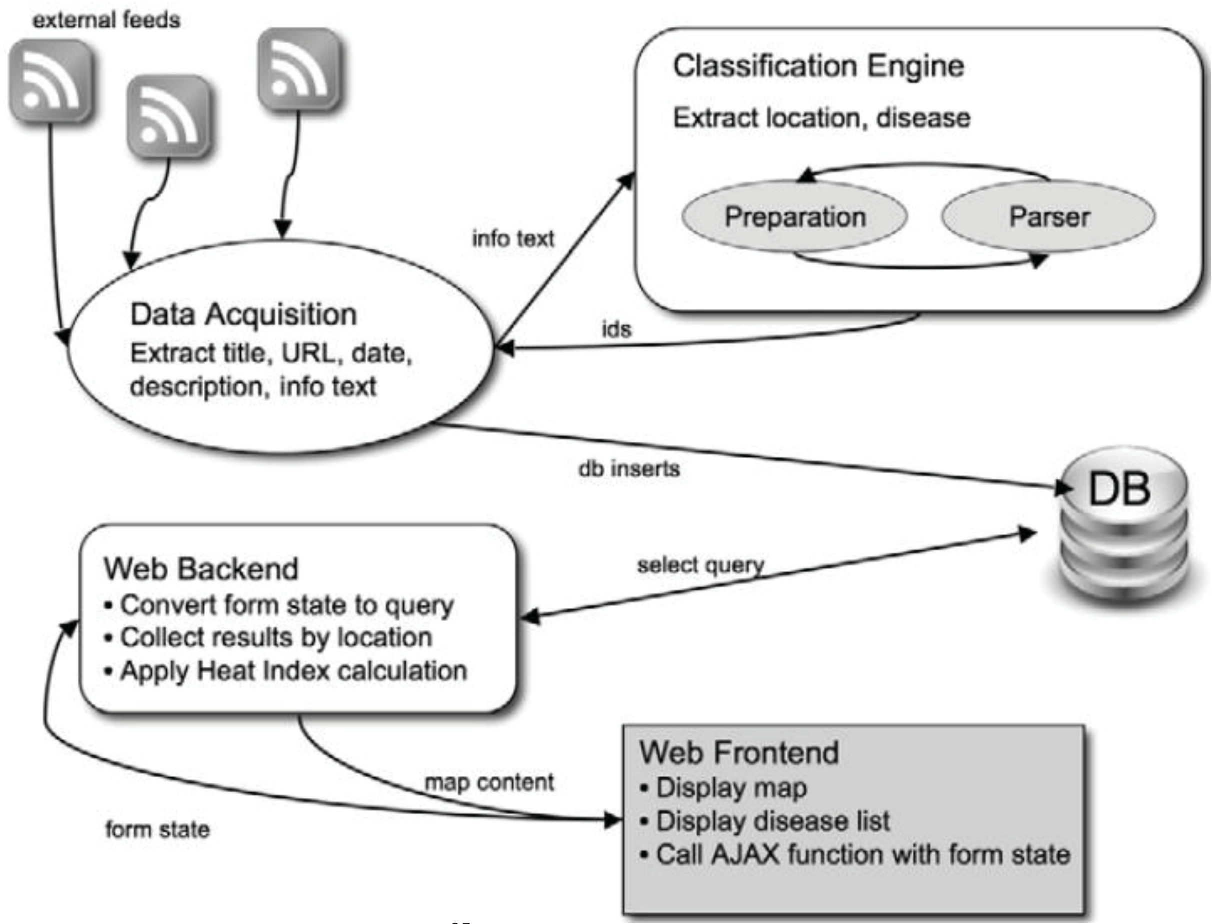

Figure 10 HealthMap System Architecture[35]

The system integrates outbreak data from a variety of electronic sources such as online news wires, RSS feeds, ProMED Mail, WHO validated official alerts and the Eurosurveillance RSS publishing which is a multi-national outbreak news site.[36] These raw data are loaded onto the Data Acquisition Engine which converts each disease outbreak report into a standard “alert” format, containing four fields: headline (the alert headline), date (the date of issue of the alert), description (brief summary of the alert), and info text (the text that will be fed into the parsing engine for the initial classification pass).[35]

Data Processing

The Classification Engine determines and classifies each alert by location and disease associated with it. It is comprised of two modules: the Preparation Module, which takes the raw input from the source, segments it and prepares it for input to the parser, and the Parser Module, which takes text input and produces disease and location codes as output.[35]

Once classified, articles are filtered by their relevance into five categories

1. **Breaking News:** newly discovered outbreak;

2. **Warning:** initial concerns of disease emergence, for example, in a conflict zone or natural disaster area;

3. **Follow-up:** reference to a past, already known outbreak;

4. **Background/Context:** information on disease context, such as vaccination campaigns, preparedness planning, research results and

5. **Not Disease Related:** information not relating to any disease or health condition.[36]

Once the alerts are classified, the system stores them in the Database.[35]

Data Dissemination

Only "breaking news" articles are overlaid on an interactive geographic map.[36,38] HealthMap currently processes 133.5 disease alerts per day on average, with approximately 50% categorised as Breaking News (65.3 reports/day).

It receives about 15,000 unique visitors per month from around the world. It is cited as a resource on the websites of organizations such as UN, FDA, Italy's national epidemiology agency (Centro Nazionale di Epidemiologia, Sorveglianza e Promozione della Salute) and many library websites and university course materials. Their most avid users tend to come from government-related domains, including the WHO, the CDC, the ECDC, US Health and Human Services and the US Department and other national, state and local bodies around the world.[36]

3.2.9 International System for Total Early Disease Detection (InSTEDD)

Table 15 Overview of InSTEDD

No	Field Name	Field Description
1	System name	International System for Total Early Disease Detection (InSTEDD)
2	System category	Moderated system
3	Country	USA
4	Year started	2006
5	Coordinating organization	
6	Purpose	To streamline the collaboration between domain experts and machine learning algorithms for detection, prediction and response to health-related events (such as disease outbreaks).
7	Jurisdiction	International (Asia, Africa, Europe and the Americas)
8	Languages	English
9	Disease type	Multiple infectious diseases
10	Public access	Restricted
11	Data processing	InSTEDD platform consists of several high-level modules, including: 1) Data gathering, 2) Automatic feature extraction, data classification and tagging, 3) Human input, hypotheses generation and review, 4) Predictions and alerts output, and 5) Field confirmation and feedback.
12	Dissemination of data	InSTEDD advises organizations like the UN, WHO and CDC on the strategic implementation of health information systems and collaboration technology ventures.
13	Most avid users	InSTEDD works with governments, universities, corporations, international health organizations, humanitarian NGOs and local communities around the world.
14	System evaluation	Not evaluated
15	Homepage	http://www.instedd.org

InSTEDD was founded in 2006 to serve as a catalyst to empower individuals and organizations by using technology for more effective action in early warning, prevention and response to disasters and public health threats.[61] InSTEDD is a hybrid (event-based and indicator-based) surveillance platform which consists of several high-level modules, including: 1) Data gathering, 2) Automatic feature extraction, data classification and tagging, 3) Human input, hypotheses generation and review, 4) Predictions and alerts output, and 5) Field confirmation and feedback.[39]

Data Acquisition

The first module is the data gathering module which allows users to collect information from several sources such as SMS messages, RSS feeds, email list like ProMed, documents, web pages, electronic medical records, animal disease data, environmental feed, remote sensing, etc.

Data Processing

The second module is the automatic feature extraction, data classification and tagging module. This module is an extensible architecture that allows the introduction of machine learning algorithms (e.g., Bayesian). These components extract and augment the features (or metadata) from multiple data streams; such as: source and target geo-location, time, route of transmission (e.g., person-to-person, waterborne), etc. In addition, these components help detect relationships between these extracted features within a collaborative space or across different collaborative spaces. The third module is human input, hypotheses generation and review. With human input, these components can suggest possible events or event types. The human input and review module is exposed as a set of functionalities that allows users to comment, tag, and rank the elements (positive, neutral, or negative). Additionally, users can generate and test multiple hypotheses in parallel, further collect and rank sets of related items (evidence), and model against baseline information (for cyclical or known events). The last module is the field confirmation and feedback. The platform maintains a list of ongoing possible threats allowing domain experts to focus their field information and either confirm or reject the hypotheses created. This feedback is then fed into the system to update (increase or decrease) the reliability of the sources and credibility of the users in light of their inferences or decisions.[39]

Data Dissemination

At the policy level, InSTEDD advises organizations like the UN, WHO and CDC on the strategic implementation of health information systems and collaboration technology ventures.[61]

3.2.10 Medical Information System (MedISys) and Pattern-based Understanding and Learning System (PULS)

Table 16 Overview of MedISys

No	Field Name	Field Description
1	System name	Medical Information System (MedISys) and Pattern-based Understanding and Learning System (PULS)
2	System category	Automatic system
3	Country	EU
4	Year started	2004
5	Coordinating organization	European Commission Directorate General *Health and Consumer Protection* (DG SANCO).
6	Purpose	To support national and international Public Health institutions in their work on monitoring health-related issues of public concern, such as outbreaks of communicable diseases, bio-terrorism, and large-scale chemical incidents.
7	Jurisdiction	International (more particularly Europe but also monitors all other continents).
8	Languages	43
9	Disease type	Multiple infectious diseases
10	Public access	Provides access at multiple levels: 1. Free public access 2. Restricted access for PH professionals outside the European Commission (EC) 3. Full access inside the EC
11	Data processing	MedISys monitors an average of 50,000 news articles per day from about 1400 news portals around the world in 43 languages and sends the document as plain text through an RSS channel to PULS which performs deeper analyses by extracting the name of the disease, the date and the location of the outbreak, the number of victims, their status and a description of the victims (i.e humans, animals, etc). PULS then returns the extracted information to MedISys in a structured form.
12	Dissemination of data	MedISys issues alert on its web page through various online statistics in deferent languages. Additionally it issues e-mail alert to its subscribers as well as SMS alerts.
13	Most avid users	WHO, CDC, ECDC, Eurosurveillance, GPHIN, **Swiss** Federal Office of Public Health, **Icelandic** Ministry of Health, **Spanish** Ministry of Sanitation, Spanish Ministry of Health and Consumer Protection, Institut de Veille Sanitaire **France, Danish** Emergency Management Agency, **Italian** Ministry of Health, Italian Ministry of Defence, **Dutch** Institute of Public Health, Dutch Food and Consumer Product Safety Authority.
14	System evaluation	Yes[41,42]
15	Homepage	http://medusa.jrc.it/medisys/homeedition/en/home.html http://sysdb.cs.helsinki.fi/tomcat/tkt_plus/jrc/ http://puls.cs.helsinki.fi/medical/

MedISys is part of the Europe Media Monitor (EMM) product family[62] developed at the European Commission's Joint Research Centre (JRC). MedISys was initiated by

European Commission Directorate General Health and Consumer Protection (DG SANCO) in 2004. Its objective is to support national and international Public Health organisations monitor issues of Public Health concern by gathering documents of relevance, filter, classify, aggregate information and alert users on topics such as outbreak of contagious diseases, nuclear and chemical incidents, bioterrorism and harmful substances such as anthrax, crowd control agents etc.[63]

Information Retrieval (IR) is an early-warning functionality of MedISys and it inter-operates with the information extraction (IE) of PULS which was developed at the University of Helsinki to extract factual information from plain text. PULS has been adapted to analyse texts in the epidemiological domain, for processing documents that triggers MedISys alerts.[64] The PULS system analyses the documents identified by MedISys, retrieves from them events, or structured facts about outbreaks of communicable disease, aggregates the events into a database, and highlights the extracted information in the text. This event extraction helps to narrow down the selection of relevant articles found in the IR step thus improves precision.[42]

Data Acquisition

MedISys monitors an average of 50,000 news articles per day from about 1400 news portals around the world in 43 languages. Some of these sources include ProMED-Mail, web sites of national public health authorities, about 150 specialist web sites (including Eurosurveillance), news from about twenty news wires, and a number of copyright-protected news sources, such as Lexis-Nexis (www.lexisnexis.com/). These monitored sources are selected strategically with the aim of covering all major European news portals, plus key news sites from around the world, in order to achieve wide geographical coverage.[8,42]

Data Processing

MedISys process the information obtained using a scraper software which automatically generates an RSS feed from these pages by means of specialised transformations and the main news article is extracted from each web page using a (patent-pending) text extraction process. MedISys can independently disseminate its results in the form of alert, but in order to perform deeper text analysis to achieve a higher precision, it works collaboration with PULS. MedISys serves as an Information Retrieval (IR) while PULS serves as an Information Extraction (IE) system.[42] IE is a text understanding technology, designed to find facts in natural language text, and transform them into a structured representation, such as a database table. The task of the PULS IE system is to find descriptions of infectious epidemics as they spread around the world.

For each disease outbreak that the system detects, it determines:

- the **name of the disease**,
- the **date** and the **location** of the outbreak,
- the **number** of victims,
- their **status** (are they sick or dead),
- a **description** of the victims (are they humans, animals, etc).[65]

There is a special RSS tunnel established between MedISys and PULS. MedISys monitors in 43 different languages and passes on the documents to PULS only in English because PULS processes only English-language documents. The documents arrive as plain text. A document batch is sent every 10 minutes, with documents newly discovered on the Web.

PULS analyses the documents, extracts incidents, and stores them in its database (http://sysdb.cs.helsinki.fi/tomcat/tkt_plus/jrc/). PULS then uses document-local heuristics to compute the confidence of the attributes in the extracted incidents. The attribute is considered confident if the score of the best value exceeds a certain threshold. Some of the attributes of an incident are considered to be more important than others. In the case of epidemic events, the principal attributes are the disease name, location and date. The system then aggregates the extracted incidents into outbreaks, across multiple documents and sources. The aggregation process requires that at least one of the incidents in each outbreak chain must be confident. Chains composed entirely of non-confident incidents are discarded. PULS then returns the extracted information to MedISys in a structured form through the tunnel in a 10 minutes interval. While both sites operate in real-time, the communication is asynchronous. PULS returns the most recent 50 incidents, filtering out duplicates: if multiple incidents of the same disease in the same location are reported, PULS returns only the most recent one. The goal of PULS is to return a set of incidents with high confidence and low redundancy.[42] The illustration MedIsys/PULS intergration can be seen in Figure 11.

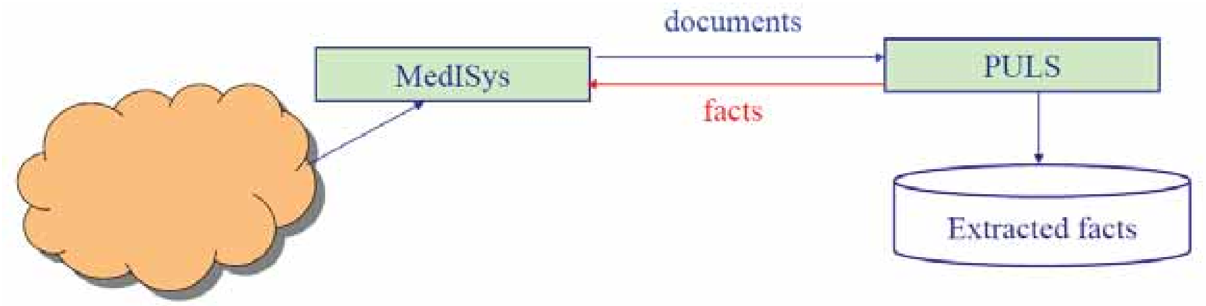

Figure 11 MedISys/PULS intergration. There is a live exchange of data between the systems every 10 minutes through an RSS channel. MedISys sends new documents to PULS. PULS extracts, then returns latest and most urgent facts to MedISys which is the disseminated.[63]

Data Dissemination

The web interface of MedISys can be used to view the latest trends and to access articles about diseases and countries. MedISys displays the title and the first few words of each article in various languages, plus a link to the URL where the full news text was originally found. The main page shows the five most recent events which corresponds to the most urgent news. For more detail, this box has a link to the batch of 50 most recent incidents. Users can subscribe to summary reports containing information on groups of alerts as well as obtain access to the JRC's Rapid News Service, (RNS), which allows them to filter news from selected sources, or countries, and which provides functionality to quickly edit and publish newsletters and to distribute them via email or to mobile phones (SMS).[42]

3.2.11 MITRE Text and Audio Processing System (MiTAP)

Table 17 Overview of MiTAP

No	Field Name	Field Description
1	System name	MITRE Text and Audio Processing System (MiTAP)
2	System category	Automatic
3	Country	USA
4	Year started	2001
5	Coordinating organization	The MITRE Corporation
6	Purpose	To provide timely, multi-lingual, global information access to analysts, medical experts and individuals involved in humanitarian assistance and relief work.
7	Jurisdiction	International
8	Languages	8 Languages (English, Chinese, French, German, Italian, Portuguese, Russian, and Spanish).
9	Disease type	Multiple infectious diseases
10	Public access	Free subscription

No	Field Name	Field Description
11	Data processing	MiTAP architecture has 3 basic components: information capture, information processing, and user interface. MiTAP captures information from sources such as web sites, electronic mailing lists, newsgroups, news feeds, and audio/video data. The data from all of these sources are then sent on to the processing phase, where the individual TIDES component technologies are employed. Entities such as person, organization and location names as well as dates, diseases, and victim descriptions are extracted. Finally, the document is processed by WebSumm which generates modified versions of extracted sentences as a summary.
12	Dissemination of data	The processed messages are converted to HTML, with colour-coded named entities, and routed to newsgroups. The newsgroups are organized by category (i.e., source, disease, region, language, person, and organization) to allow analysts, with specific information needs, to locate material quickly.
13	Most avid users	American Red Cross, the United Nations, and the European Disaster Center, medical analysts, doctors, government and military officials, members of non-governmental organizations, and members of humanitarian assistance/disaster relief organizations.
14	System evaluation	Yes[43]
15	Homepage	http://www.mitre.org/

The MiTAP (MITRE Text and Audio Processing) system was developed as an experimental prototype using human language technologies for monitoring disease outbreaks and other global events. It was created to explore the integration of synergistic TIDES language processing technologies: Translation, Information Detection, Extraction, and Summarization.

Multiple information sources such as epidemiological reports, newswire feeds, email, online news in multiple languages are automatically captured, filtered, translated, summarized, and categorized into searchable newsgroups based on disease, region, information source, person, organization, and language.

MiTAP is designed to provide the end user with timely, accurate, novel information and present it in a way that allows the analyst to spend more time on analysis and less time on finding, translating, distilling and presenting information.

Figure 12 illustrates the three phases of the underlying MiTAP architecture: information capture, information processing, and user interface.

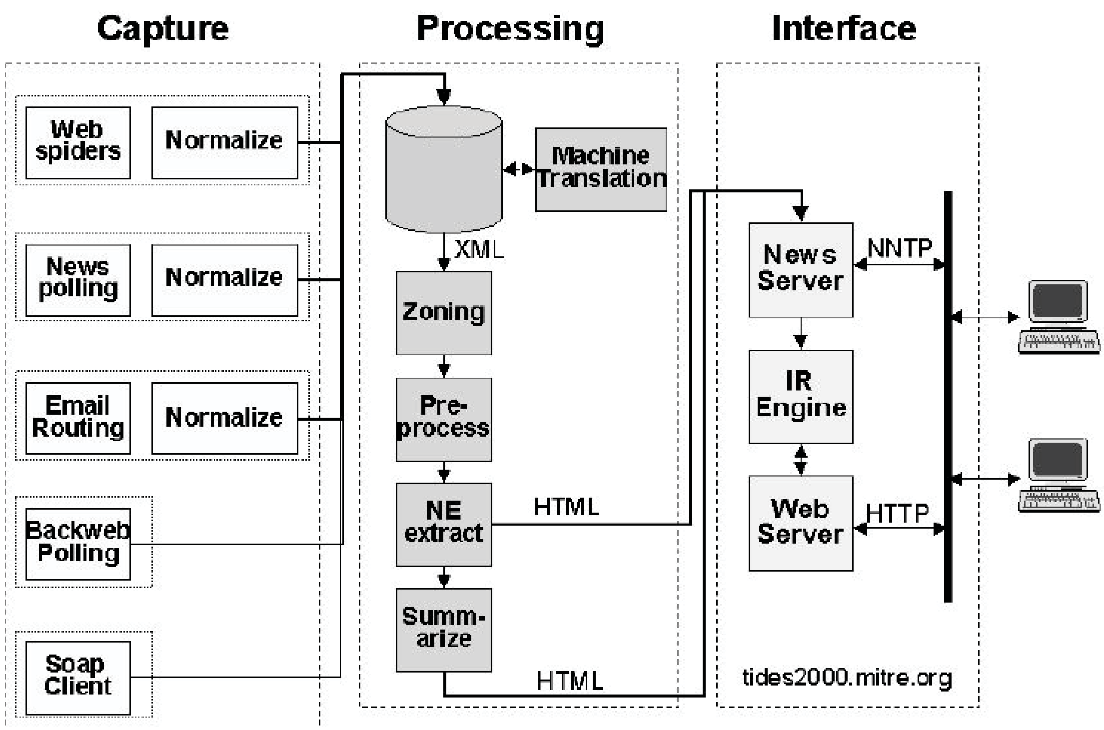

Figure 12 MiTAP architecture[43]

Data Acquisition

The capture process supports web sources, electronic mailing lists, newsgroups, news feeds, and audio/video data. The first four of these categories are automatically harvested and filtered, and the resulting information is normalized prior to processing. A ViTAP system captures and transcribes TV news broadcasts, making the text transcriptions available to MiTAP. The data from all of these sources are then sent on to the processing phase, where the individual TIDES component technologies are employed.

Data Processing

Each normalized message is passed through a zoner that uses human-generated rules to identify the source, date, and other fields such as headline or title, article body, etc. The zoned messages are pre-processed to identify paragraph, sentence, and word boundaries as well as part-of-speech tags. The pre-processed messages are then fed into a named entity recognizer (a natural language analyzer), which identifies person, organization and location names as well as dates, diseases, and victim descriptions using human-generated rules. Finally, the document is processed by WebSumm[66] which generates modified versions of extracted sentences as a summary. For non-English sources, a machine translation system is used to translate the messages automatically into English. The final phase is the user interface.

Data Dissemination

The processed messages are converted to HTML, with colour-coded named entities, and routed to newsgroups. The newsgroups are organized by category (i.e., source, disease, region, language, person, and organization) to allow analysts, with specific information needs, to locate material quickly. Users can access MiTAP through any standard newsreader and customize their view by subscribing to specific newsgroup categories based on their needs.

MiTAP has incorporated two types of multi-document summarization into the system. Newsblaster (http://www.cs.columbia.edu/nlp/newsblaster) automatically clusters articles and generates summaries based on each cluster and Alias I (http://www.alias-i.com) which generates daily top ten lists of diseases in the news. These summaries are posted to the MiTAP news server on a daily basis and act as both high-level views of the day's news and as points of entry into the system. Newsblaster automatically summarizes clusters of documents. Users can access the complete MiTAP articles for context or further reading.

MiTAP currently stores over one million articles and processes additional 2000 to 10,000 daily, delivering up-to-date information to dozens of regular users.[43]

3.2.12 Program for Monitoring Emerging Diseases (ProMED)-mail

Table 18 Overview of ProMED-mail (PMM)

No	Field Name	Field Description
1	System name	Program for Monitoring Emerging Diseases (ProMED)-mail
2	System category	Moderated
3	Country	USA
4	Year started	1994
5	Coordinating organization	International Society for Infectious Diseases (ISID)
6	Purpose	To promote communication amongst the international infectious disease community, including scientists, physicians, epidemiologists, public health professionals, and others interested in infectious diseases on a global scale.
7	Jurisdiction	International (about 185 countries with more than 40,000 subscribers).

No	Field Name	Field Description
8	Languages	7 (English, Portuguese [ProMED-PORT], Spanish [ProMED-ESP], Russian [ProMED-RUS] French [ProMED-FRA], Chinese and Japanese).
9	Disease type	Multiple infectious diseases and toxins, and generally excludes reports of HIV:AIDS and tuberculosis.
10	Public access	Free subscription
11	Data processing	PMM receives information from subscribers and other sources such as raw newspaper articles, TV segment, or radio reports in addition to the web sites of ministries of health; federal, state and local health departments; and international organizations The top moderator filters all incoming information. Most reports are examined carefully and then sent to one of the expert subject moderators for further review. The specialty moderator then assesses the reliability and accuracy of the information which may involve verification of the reports. Edited reports are returned to the top moderator for final editing, verification, and additional commentary. ProMED reports are coded Green (for normal reports), Yellow (for urgent reports) or Red (extremely urgent) based on priority.
12	Dissemination of data	Reports are distributed by email to direct subscribers and posted immediately on the ProMED-mail web site.
13	Most avid users	WHO, CDC, UN humanitarian and relief agencies, Federation of International Red Cross and Red Crescent Societies, Laboratory Centers for Disease Control, Canada, Public Health Laboratory Service UK, Pasteur Institutes in France, Tahiti, Vietnam, National Institute of Health, Japan and national health ministries and departments worldwide, as well as thousands of interested members of the general public.
14	System evaluation	Yes[67]
15	Homepage	http://www.promedmail.org

In 1994 ProMED-mail (PMM) was formed as an initiative of the Federation of American Scientists (FAS), with technical support from SatelLife of Boston, Massachusetts, USA and In 1999 PMM left FAS and SatelLife to become a program of the International Society for Infectious Diseases (ISID), in collaboration with the Harvard School of Public Health and the Oracle Corporation.[45] It was established to provide an early global warning of emerging diseases of humans, animals, and plants as well as of disease activities signalling biological warfare and bioterrorist activities.[44] The central purpose is to promote communication amongst the international infectious disease community, including scientists, physicians, epidemiologists, public health professionals, and others interested in infectious diseases on a global scale.[68]

PMM concentrates on reports of outbreaks of infectious diseases and toxins, and generally excludes reports of HIV:AIDS and tuberculosis, which have their own reporting systems.[46] PMM is open to all sources and free of political constraints. It currently has over 40,000 subscribers in at least 185 countries.[68]

It disseminates it results in 7 languages including English, French, Spanish, Portuguese, Chinese, Russian and Japanese. ProMED-mail is in operation 7 days a week, 365 days a year and the system information flowchart can be seen in Figure 13.

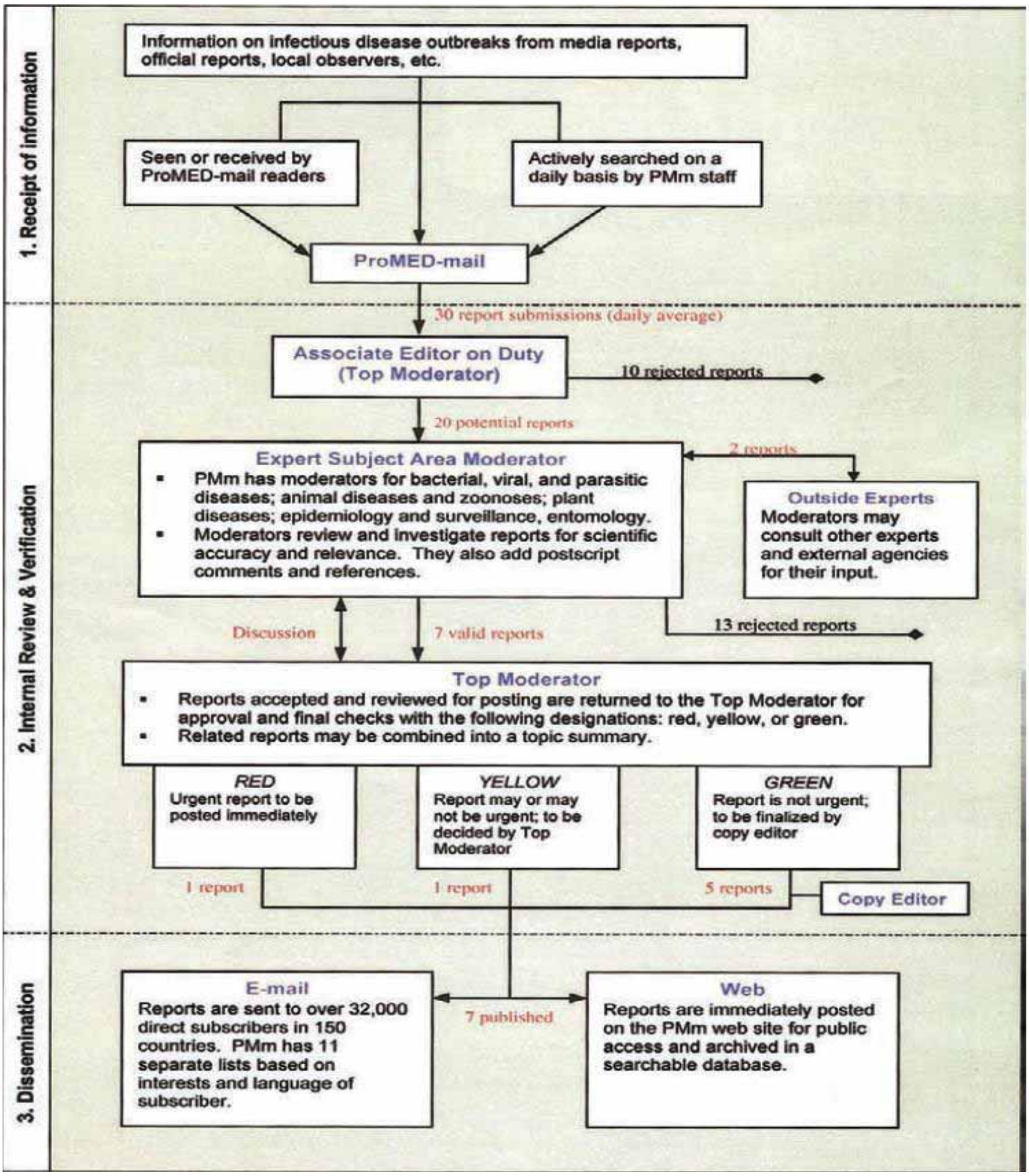

Figure 13 Promed-mail information flow[48]

Data Acquisition

PMM receives dozens of e-mails, many from subscribers, that contain new data on outbreaks or diseases, some of which are reported firsthand[48] and some of which are reported from other sources such as raw newspaper articles, TV Segment, or radio reports.[44] In addition, ProMED staff search through the Internet and traditional media for relevant items and scans a variety of official and unofficial web sites such as the web sites of ministries of health; federal, state and local health departments; and international organizations looking for recent updates.[48]

Data Processing

The top moderator (i.e. either the editor or one of the associate editors depending on who is on duty) filters all incoming information. Reports which contain information that is irrelevant, not credible, or outdated or duplicates information contained in previous reports are immediately rejected. Most reports are examined carefully and then sent to one of the expert subject moderators (experts in viral diseases, bacterial diseases, plant diseases, veterinary diseases and zoonoses, epidemiology, and medical entomology) for further review. Results are frequently sent to more than one moderator. Some reports could be sent for translation to outside experts for their opinions. The specialty moderator then has several tasks which are to assess the reliability and accuracy of the information which may involve verification of the report from another source, including direct contact with a colleague who might possess first-hand knowledge.

The specialty moderator edits the piece for content, provides pertinent references and adds brief commentary with the intention of providing background and perspective. Edited reports are returned to the top moderator for final editing, verification, and additional commentary.

ProMED reports are coded Green, Yellow or Red based on priority. Reports of normal priority (green) are handled routinely and will normally transit the system, from initial report to final posting. Reports of greater urgency (Yellow) for example, a report of an outbreak of a potentially epidemic disease in a new location receive expedited review and are posted as quickly as possible. Reports that are extremely urgent (red) may bypass parts of the editorial process to be posted immediately. The entire process occurs in less than 24 hours.[48]

Data Dissemination

Finalized reports are simultaneously posted to the PMM web site and distributed to mailing lists that are based on the interests and languages of the subscribers.[48] PMM posts an average of seven reports each day, 365 days per year.[50]

3.2.13 Proteus-BIO

Table 19 Overview of Proteus-BIO

No	Field Name	Field Description
1	System name	Proteus-BIO
2	System category	Automatic
3	Country	USA
4	Year started	
5	Coordinating organization	Computer Science Department New York University
6	Purpose	To permit the monitoring of a large number of news sources for reports of infectious disease outbreaks around the world.
7	Jurisdiction	International
8	Languages	English
9	Disease type	Multiple infectious diseases
10	Public access	Freely available
11	Data processing	Proteus-BIO consists of a 3 basic components: a web crawler, an extraction engine and a database browser. The web crawler gathers relevant documents from websites and local newspapers; the extraction engine analyzes the text of the story and identifies instances of infectious disease outbreaks such as the location and date of the outbreak, the disease, the number and type of victims and their status(i.e. if they died or not). The information extraction engine converts the individual outbreak events to a tabular database.
12	Dissemination of data	The database browser provides users access to the events and, through them, to the documents.
13	Most avid users	
14	System evaluation	Yes[52]
15	Homepage	

Proteus-BIO Prototype event-based system is managed by the Computer Science Department, New York University. The purpose of the system is to permit the monitoring of a large number of news sources for reports of infectious disease outbreaks around the world. The system architecture can be seen in Figure 14.

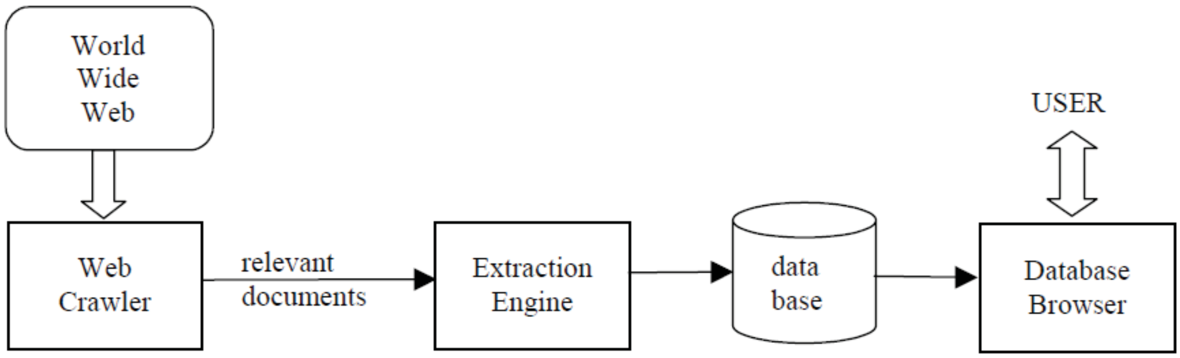

Figure 14 System architecture of Proteus-BIO[52]

Data Acquisition

The Web crawler traverses portions of the web each night, looking for new, relevant web pages. It searches selected general and medical news sources, looking for new web pages. It visits two medical news sites which are ProMed-mail (http://www.promedmail.org) and the Disease Outbreak News of the WHO (http://www.who.int/csr/don/en/). Proteus-Bio also searches other local newspaper sources such as the Chicago Sun-Times, and one general news search engine, Northern Light.

In the case of the medical news sources, all new web pages are sent to the extraction engine. For general news, a simple filter is applied which requires that a document contains at least one word or phrase relevant to an infectious disease outbreak. This filter is intended only to improve efficiency.

The web crawler also finds the text body within the web page because a typical web page has lots of information besides the actual text of a story such as headlines, links to other stories, sponsorship information or advertisements, etc. For most web pages, the crawler uses the HTML markup to locate the relevant text. For the ProMed web pages, which contain primarily text without HTML markup, the crawler uses specific text tags and other layout indicators (blank lines, capitalized lines, etc.).

Data Processing

The extraction engine analyzes the text of the story and identifies instances of infectious disease outbreaks. For each outbreak report, it captures specific pieces of information: the location and date of the outbreak, the disease, the number and type of victims, and whether they died. For each such report, the engine adds one row to the database of outbreaks, and links the row back to the document. The engine operates by looking for linguistic patterns in the text, such as “outbreak of <disease> killed <victims>”, and uses the variables in the pattern, such as <disease> and <victims>, to fill slots in the database.

Data Dissemination

The database browser provides the user interface for the system. It presents the extracted information in tabular form and each incident (database entry) is presented as a separate row in the table. Each row is linked back to the corresponding passage in the document, and the relevant items (the disease, the date, the location, the victim description) are highlighted and colour-coded.[52]

4. Discussion and Recommendations

The purpose of this systematic literature review was to find out which event-based surveillance systems exist, where they are based and which systems have been evaluated. We present for the first time the list of all event-based public health surveillance systems that exist till date. We identified 13 systems for which 2 systems (MiTAP and Proteus-BIO) are prototype systems and 10 of the 13 systems have been evaluated. As seen in Figure 4 (page 18), approximately 70 % of the event-based systems are based in the US, followed by EU (as a whole) with about 15 % and lastly Canada and Japan with 8 % each. There was no event-based system found in all other countries grouped together. Figure 5 (page 18) illustrates that North America is the leading continent with 77% of event-based surveillance systems. There is a wide difference with Europe which comes second with 15 % of the event-based systems and Asia comes next with 8%. No event-based system was found in Africa, Australia and South America.

10 systems (77 %) were found to have been evaluated and these include; Argus, BioCaster, EWRS, GOARN, GPHIN, HealthMap, MedISys, MiTAP, ProMED and Proteus-BIO. We found no evaluation on 3 systems (23%) which include; EpiSPIDER, GODS[N] and InSTEDD. A system was considered evaluated if it were stated in the study or if it were stated in some other study. The need for seeking evaluation of public health surveillance systems is to ensure that problems of public health importance are being monitored efficiently and effectively.[3]

Our results show that while North America and Europe are leading in event-based surveillance systems Africa, Asia, Australia and South America posses little or no event-based systems to monitor their epidemic threats. A study which analyzed the entire WHO public record of *Disease Outbreak News* reports and created a catalogue of selected WHO confirmed outbreaks that occurred during 1996–2009[69] showed that Africa, Asia and South America are the 3 continents having the most number of emerging infectious disease events. The common denominator between these three most affected continents is the high density of under-developed countries in them. There are a number of reasons centred on the developmental status of the country in question which includes; a variety of socio-economic, health, sanitation, and settlement patterns. Under-developed countries typically have a poor or insufficient health-care infrastructure and are therefore unable to cope with a large disease outbreak.[70] The early detection and prevention of infectious disease at the early stage of an event can help prevent an epidemic outbreak or reduce its impact. Modern technologies, mainly related to development of the internet, have rapidly changed the way we access health information[60] and surveillance institutions have recently been actively searching for information about health threats using internet scanning tools, email distribution lists or networks that complement the early warning function of routine surveillance systems. [71] These online media, scientific forums and direct electronic communication now allow us to shortcut traditional reporting mechanisms that travel through the various levels of public health administration[60] and thus allow us to detect disease outbreaks earlier with reduced cost and increased reporting transparency.[4]

To move forward with this type of work, the sub-areas of Infoveillance and Infodemology[72] may provide a way to examine the science behind event-based systems. Infodemiology can be defined as the science of distribution and determinants of information in an electronic medium, specifically the Internet, or in a population, with the ultimate aim to inform public health and public policy. Potential infodemiology indicators and metrics include automatically aggregated and analyzed data on the prevalence and patterns of information on websites and social media; metrics on the “chatter” in discussion groups, blogs, and microblogs (eg, Twitter); and activities on search engines, etc. Infoveillance has the same definition as infodemiology but with a primary aim of surveillance, and hence the name infoveillance.

Presently, indicator-based surveillance is being complemented by event-based surveillance; we know this because our results show 13 event-based systems that are currently being used. Since event-based component of epidemic intelligence are relatively cheaper to set up than the indicator-based component, under-developed countries should therefore use the advantage of the increase use of internet and the increased amount of information that can be obtained from the social media to set up their event-based systems which will complement their indicator-based systems. Public health officials from under-developed countries could also subscribe to those event-based systems which monitor over their jurisdiction and obtain early warning signals which will enable them to take rapid intervention in due time. This review provides a useful report that is up to date with what is current in the field of event-based surveillance and may provide the background to public health officials in developing new event-based systems or enhancing their indicator-based work.

5. Conclusion

This systematic literature review indicates that there are currently 13 event-based surveillance systems being used with North America being the leading continent where the systems are based, followed by Europe. Africa, Asia and Latin America which are the 3 most affected continents with emerging infectious diseases posses little or no event-based systems to monitor their epidemic threats. Under-developed countries should take the advantage in the advancement of modern information technology to set up their respective event-based systems which could be relatively cheap but effective to complement their indicator-based systems or subscribe to the already existing systems which monitor over their jurisdictions. This review may therefore provide the necessary background to public health officials in developing new event-based systems or enhancing their indicator-based work.

6. Limitations

Our systematic review has 5 potential limitations. First, since the scope of our review was to focus only on event-based systems while excluding bioterrorist related systems, we may therefore have neglected to include some potential relevant event-based surveillance systems which were referred to as bioterrorist related systems.

Second, we might also have excluded 5 potential systems (Appendix 3; Table 22) which were initially considered as event-based systems after reading their abstracts (Figure 3; Page 10), but were subsequently eliminated after accessing their full texts due to insufficient information to classify the systems i.e. many details of the features of the systems were not readily available from the published information about these systems.

Third, data on some existing event-based surveillance systems may not be publicly available and hence we might not have been able to pickup such event-based systems. Examples of such systems include systems developed by military or public health officials whose objective it is to deploy and maintain surveillance for detecting outbreaks in their jurisdiction but whose mandate does not necessarily include publishing.

Fourth, our review focused mainly on published literature in 3 databases (Pubmed, Scopus and Scirus) and some papers describing event-based surveillance systems might be in the realm of “grey” or unpublished literature.

Fifth, one exclusion criterion was to eliminate articles with no available abstracts; we might therefore have eliminated some potential event-based systems published with no abstract.

Appendix 1

Table 20 Data extracted from articles

No	Field Name	Field Description
1	System name	The name of the system
2	System category	Event detection systems can be classified into one of these three categories:[8] • News aggregators • Automatic systems • Moderated systems
3	Country	Country where the system was founded
4	Year started	The year the system started operating
5	Coordinating organization	The unit which operates the system
6	Purpose	The purpose of the system
7	Jurisdiction	The location of the system and their highest level of aggregation. i.e if the system operates both at the local and national level, then it would be entered as national and for those that operate in more than one country, they will be entered as international.
8	Languages	The number of languages the system covers or gets information from
9	Disease type	Type of diseases covered by the system; more than three will be labelled as “multiple infectious disease”
10	Public access	If it is freely accessible to the general public or restricted only to registered members
11	Data processing	A description of the methods employed by the system to collect the necessary data, and how the data are analysed
12	Dissemination of data	How the system disseminates its data
13	Most avid users	The organizations or individuals who obtain information from the event-based system
14	System evaluation	If the article mentioned that system was evaluated or if it was mentioned in a different article.
15	Homepage	The system’s homepage

Appendix 2

Table 21 List of all 32 articles describing event-based surveillance systems

System	Number	Reference number	Bibliographic citations
Argus	1	21	James M, Wison V. Argus: A Global Detection and Tracking System for Biological Events. 2007. Available at: http://www.isdsjournal.org/article/viewArticle/1953 . Accessed November 29, 2010.
	2	22	Nelson NP, Brownstein JS, Hartley DM. Event-based biosurveillance of respiratory disease in Mexico, 2007-2009: connection to the 2009 influenza A(H1N1) pandemic? *Euro Surveill*. 2010;15(30). Available at: http://www.ncbi.nlm.nih.gov/pubmed/20684815. Accessed April 8, 2011.
Biocaster	3	23	Collier N, Doan S, Kawazoe A, et al. BioCaster: Detecting public health rumors with a Web-based text mining system. *Bioinformatics*. 2008;24(24):2940-2941.
EpiSPIDER	4	24	Tolentino H, Kamadjeu R, Fontelo P, et al. Scanning the Emerging Infectious Diseases Horizon - Visualizing ProMed Emails Using EpiSPIDER. 2007. Available at: http://www.lhncbc.nlm.nih.gov/lhc/docs/published/2007/pub2007055.pdf. Accessed November 29, 2010.
	5	25	Keller M, Blench M, Tolentino H, et al. Use of unstructured event-based reports for global infectious disease surveillance. *Emerging Infect. Dis*. 2009;15(5):689-695.
EWRS	6	26	Guglielmetti P, Coulombier D, Thinus G, Van Loock F, Schreck S. The early warning and response system for communicable diseases in the EU: an overview from 1999 to 2005. *Euro Surveill*. 2006;11(12):215-220.
GOARN	7	27	Heymann DL, Rodier GR. Hot spots in a wired world: WHO surveillance of emerging and re-emerging infectious diseases. *The Lancet Infectious Diseases*. 2001;1(5):345-353.
GODS[N]	8	28	Khan SA, Patel CO, Kukafka R. GODSN: Global News Driven Disease Outbreak and Surveillance. *AMIA Annu Symp Proc*. 2006:983.
GPHIN	9	29	Mawudeku A, Blench M. Global Public Health Intelligence Network (GPHIN). *Proceeding of the 7th Conference of the Association for Machine Translation in the Americas Cambridge, Massachusetts, United States of America 2006, 7-11*. 2006. Available at: http://www.mt-archive.info/MTS-2005-Mawudeku.pdf.
	10	30	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;97(1):42-44.

System	Number	Reference number	Bibliographic citations
HealthMap	11	31	Keller M, Freifeld CC, Brownstein JS. Automated vocabulary discovery for geo-parsing online epidemic intelligence. *BMC Bioinformatics*. 2009;10:385.
	12	32	Brownstein JS, Freifeld CC. Evaluation of Internet-Based Informal Surveillance for Global Infectious Disease Intelligence. *International Journal of Infectious Diseases*. 2008;12(Supplement 1):e193-e194.
	13	33	Brownstein JS, Freifeld CC, Reis BY, Mandl KD. Evaluation of online media reports for global infectious disease intelligence. *Advanced in disease surveillance 2007;4:3*. 2007. Available at: http://www.isdsjournal.org/articles/1935.pdf. Accessed November 22, 2010.
	14	34	Chen H, Yan P, Zeng D, et al. HealthMap. In: *Infectious Disease Informatics*.Vol 21. Integrated Series in Information Systems. Springer US; 2010:183-186. Available at: http://dx.doi.org/10.1007/978-1-4419-1278-7_14.
	15	35	Freifeld CC, Mandl KD, Reis BY, Brownstein JS. HealthMap: Global Infectious Disease Monitoring through Automated Classification and Visualization of Internet Media Reports. *Journal of the American Medical Informatics Association*. 2008;15(2):150-157.
	16	36	Brownstein JS, Freifeld CC. HealthMap: the development of automated real-time internet surveillance for epidemic intelligence. *Euro Surveill*. 2007;12(11):E071129.5.
	17	37	Nelson R. HealthMap: the future of infectious diseases surveillance? *The Lancet Infectious Diseases*. 2008;8(10):596.
	18	38	Brownstein JS, Freifeld CC, Reis BY, Mandl KD. Surveillance Sans Frontières: Internet-based emerging infectious disease intelligence and the HealthMap project. *PLoS Med*. 2008;5(7):e151.
InSTEDD	19	39	Kass-Hout TA, di Tada N. International System for Total Early Disease Detection (InSTEDD) Platform. 2009. Available at: http://www.isdsjournal.org/article/viewArticle/3308 . Accessed November 19, 2010.
MedISys	20	40	Yangarber R, Steinberger R, Best C, et al. Combining Information Retrieval and Information Extraction for Medical Intelligence. *Proceeding of Mining Massive Data Sets for Security, NATO Advanced Study Institute. Gazzada, Italy 2007*. 2007. Available at: http://langtech.jrc.it/mmdss2007/htdocs/Presentations/Docs/MMDSS_Yangarber_Steinberger_PUBLIC.pdf.
	21	41	Rortais A, Belyaeva J, Gemo M, van der Goot E, Linge JP. MedISys: An early-warning system for the detection of (re-)emerging food- and feed-borne hazards. *Food Research International*. 2010;43(5):1553-1556.

System	Number	Reference number	Bibliographic citations
	22	42	STEINBERGER R, FUART F, van der GOOT E, et al. Text Mining from the Web for Medical Intelligence. 2008. Available at: http://langtech.jrc.it/Documents/2009_MMDSS_Medical-Intelligence.pdf. Accessed November 19, 2010.
MiTAP	23	43	Damianos L, Ponte J, Wohlever S, et al. MiTAP for Biosecurity: A Case Study. *AI Magazine*. 2002;23(4):13-29.
ProMED	24	44	Hugh-Jones M. Global awareness of disease outbreaks: The experience of ProMED-mail. *Public Health Reports*. 2001;116(SUPPL. 2):27-31.
	25	45	Woodall JP. Global surveillance of emerging diseases: the ProMED-mail perspective. *Cad Saude Publica*. 2001;17 Suppl:147-154.
	26	46	Woodall J. Official versus unofficial outbreak reporting through the Internet. *Int J Med Inform*. 1997;47(1-2):31-34.
	27	47	Morse SS, Rosenberg BH, Woodall J. ProMED global monitoring of emerging diseases: Design for a demonstration progam. *Health Policy*. 1996;38(3):135-153.
	28	48	Madoff LC. ProMED-mail: an early warning system for emerging diseases. *Clin. Infect. Dis*. 2004;39(2):227-232.
	29	49	Woodall J. Stalking the next epidemic: ProMED tracks emerging diseases. *Public Health Reports*. 1997;112(1):78-82.
	30	50	Madoff LC, Woodall JP. The internet and the global monitoring of emerging diseases: lessons from the first 10 years of ProMED-mail. *Arch. Med. Res*. 2005;36(6):724-730.
	31	51	Zeldenrust ME, Rahamat-Langendoen JC, Postma MJ, van Vliet JA. The value of ProMED-mail for the Early Warning Committee in the Netherlands: more specific approach recommended. *Euro Surveill*. 2008;13(6). Available at: http://www.ncbi.nlm.nih.gov/pubmed/18445424. Accessed November 18, 2010.
Proteus-BIO	32	52	Grishman R, Huttunen S, Yangarber R. Information extraction for enhanced access to disease outbreak reports. *J Biomed Inform*. 2002;35(4):236-246.

Appendix 3

Table 22 Five potential event-based systems eliminated due to insufficient information

System	Number	Bibliographic citations
Google Flu Trends	1	Anon. Google Flu Trends includes 14 European countries. *Euro surveillance : bulletin européen sur les maladies transmissibles = European communicable disease bulletin*. 2009;14(40). Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-77950307140&partnerID=40&md5=a16b343cf4aec3f084a6364ff8d7d79c. Accessed May 5, 2011.
	2	Carneiro HA, Mylonakis E. Google trends: A web-based tool for real-time surveillance of disease outbreaks. *Clinical Infectious Diseases*. 2009;49(10):1557-1564.
	3	Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data. *Nature*. 2009;457(7232):1012-1014.
	4	Mikler AR, Singh KP, Cook DJ, Corley CD. Monitoring Influenza Trends through Mining Social Media. 2009;vol.1:1-7.
	5	Pelat C, Turbelin C, Bar-Hen A, Flahault A, Valleron A-J. More Diseases Tracked by Using Google Trends. *Emerg Infect Dis*. 2009;15(8):1327-1328.
	6	Valdivia A, López-Alcalde J, Vicente M, et al. Monitoring influenza activity in Europe with Google Flu Trends: Comparison with the findings of sentinel physician networks - results for 2009-10. *Eurosurveillance*. 2010;15(29). Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-77955123105&partnerID=40&md5=601a10d7b0823a2afe4ad91408a78c7f. Accessed May 6, 2011.
	7	Watts G. Google watches over flu. *BMJ*. 2008;337(dec31 1):a3076-a3076.
	8	Wilson N, Mason K, Tobias M, et al. Interpreting Google flu trends data for pandemic H1N1 influenza: the New Zealand experience. *Euro Surveill*. 2009;14(44). Available at: http://www.ncbi.nlm.nih.gov/pubmed/19941777. Accessed April 15, 2011.
Gripenet	9	van Noort SP, Muehlen M, Rebelo de Andrade H, et al. Gripenet: an internet-based system to monitor influenza-like illness uniformly across Europe. *Euro Surveill*. 2007;12(7):E5-6.
Canadian Geospatial Data Infrastructure (CGDI),	10	Gao S, Mioc D, Yi X, et al. The Canadian geospatial data infrastructure and health mapping. *CyberGeo*. 2008;2008. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0 67650911596&partnerID=40&md5=e14b2fd4430fb074c6586c3fc0e12f83. Accessed May 4, 2011.

System	Number	Bibliographic citations
Quarantine Activity Reporting System (QARS)	11	Krishnamurthy R, Remis M, Brooke L, et al. Quarantine Activity Reporting System (QARS). *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*. 2006:990.
A Mobile-phone based system	12	Lin Y, Heffernan C. Accessible and inexpensive tools for global HPAI surveillance: A mobile-phone based system. *Prev. Vet. Med*. 2011;98(2-3):209-214.

References

1. Teutsch SM, Thacker SB. Planning a public health surveillance system. *Epidemiol Bull*. 1995;16(1):1-6.

2. Teutsch SM, Churchill RE. *Principles and Practice of Public Health Surveillance*. 2nd ed. Oxford University Press, USA; 2000.

3. German RR, Lee LM, Horan JM, et al. Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. *MMWR Recomm Rep*. 2001;50(RR-13):1-35; quiz CE1-7.

4. Wilson K, Brownstein JS. Early detection of disease outbreaks using the Internet. *Canadian Medical Association Journal*. 2009;180(8):829 -831.

5. Chanlekha H, Kawazoe A, Collier N. A framework for enhancing spatial and temporal granularity in report-based health surveillance systems. *BMC Medical Informatics and Decision Making*. 2010;10(1):1.

6. Vrbova L, Stephen C, Kasman N, et al. Systematic Review of Surveillance Systems for Emerging Zoonoses. *Transboundary and Emerging Diseases*. 2010;57(3):154-161.

7. Vrbova L, Stephen C, Kasman N, et al. Systematic Review of Surveillance Systems for Emerging Zoonoses. *Transboundary and Emerging Diseases*. 2010;57(3):154-161.

8. Linge JP, Steinberger R, Weber TP, et al. Internet surveillance systems for early alerting of health threats. *Euro Surveill*. 2009;14(13). Available at: http://www.ncbi.nlm.nih.gov/pubmed/19341610. Accessed August 13, 2011.

9. Paquet C, Coulombier D, Kaiser R, Ciotti M. Epidemic intelligence: a new framework for strengthening disease surveillance in Europe. *Euro Surveill*. 2006;11(12):212-214.

10. European Centre for Disease Prevention and Control (ECDC) - Health Comunication Unit - Eurosurveillance editorial team. The surveillance of communicable diseases in the European Union – a long-term strategy (2008-2013). 2008. Available at: http://www.eurosurveillance.org/viewarticle.aspx?articleid=18912. Accessed June 14, 2011.

11. Paquet C, Coulombier D, Kaiser R, Ciotti M. Epidemic intelligence: a new framework for strengthening disease surveillance in Europe. *Euro Surveill*. 2006;11(12):212-214.

12. Anon. A Guide to Establishing Event-based Surveillance. *WHO Library Cataloguing in Publication Data*. 2008. Available at: http://www.wpro.who.int/internet/resources.ashx/CSR/Publications/eventbasedsurv.pdf. Accessed November 22, 2010.

13. European Centre for Disease Prevention and Control (ECDC) - Health Comunication Unit - Eurosurveillance editorial team. What is epidemic intelligence,

and how is it being improved in Europe? 2006. Available at: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=2892. Accessed May 16, 2011.

14. Anon. Medical EcoSystem (M-Eco). Available at: http://www.meco-project.eu/. Accessed September 8, 2011.

15. Bravata DM, McDonald KM, Smith WM, et al. Systematic review: surveillance systems for early detection of bioterrorism-related diseases. *Ann. Intern. Med.* 2004;140(11):910-922.

16. Anon. Entrez-PubMed. Available at: http://www.nslij-genetics.org/search_pubmed.html. Accessed July 10, 2011.

17. Anon. Scopus - Elsevier. Available at: http://www.elsevier.com/wps/find/electronicproductdescription.cws_home/704746/description#description. Accessed July 10, 2011.

18. Anon. Scirus - for scientific information. Available at: http://www.scirus.com/. Accessed July 10, 2011.

19. Pramataris K, Doukidis G, Giaglis G, Raptakis J. The EUROMEDIES EDI prototype system. *Studies in health technology and informatics*. 1996;28:62-68.

20. Anon. Zotero | Home. Available at: http://www.zotero.org/. Accessed August 10, 2011.

21. James M, Wison V. Argus: A Global Detection and Tracking System for Biological Events. 2007. Available at: http://www.isdsjournal.org/article/viewArticle/1953. Accessed November 29, 2010.

22. Nelson NP, Brownstein JS, Hartley DM. Event-based biosurveillance of respiratory disease in Mexico, 2007-2009: connection to the 2009 influenza A(H1N1) pandemic? *Euro Surveill*. 2010;15(30). Available at: http://www.ncbi.nlm.nih.gov/pubmed/20684815. Accessed April 8, 2011.

23. Collier N, Doan S, Kawazoe A, et al. BioCaster: Detecting public health rumors with a Web-based text mining system. *Bioinformatics*. 2008;24(24):2940-2941.

24. Tolentino H, Kamadjeu R, Fontelo P, et al. Scanning the Emerging Infectious Diseases Horizon - Visualizing ProMed Emails Using EpiSPIDER. 2007. Available at: http://www.lhncbc.nlm.nih.gov/lhc/docs/published/2007/pub2007055.pdf. Accessed November 29, 2010.

25. Keller M, Blench M, Tolentino H, et al. Use of unstructured event-based reports for global infectious disease surveillance. *Emerging Infect. Dis*. 2009;15(5):689-695.

26. Guglielmetti P, Coulombier D, Thinus G, Van Loock F, Schreck S. The early warning and response system for communicable diseases in the EU: an overview from 1999 to 2005. *Euro Surveill*. 2006;11(12):215-220.

27. Heymann DL, Rodier GR. Hot spots in a wired world: WHO surveillance of emerging and re-emerging infectious diseases. *The Lancet Infectious Diseases*. 2001;1(5):345-353.

28. Khan SA, Patel CO, Kukafka R. GODSN: Global News Driven Disease Outbreak and Surveillance. *AMIA Annu Symp Proc*. 2006:983.

29. Mawudeku A, Blench M. Global Public Health Intelligence Network (GPHIN). *Proceeding of the 7th Conference of the Association for Machine Translation in the Americas Cambridge, Massachusetts, United States of America 2006, 7-11*. 2006. Available at: http://www.mt-archive.info/MTS-2005-Mawudeku.pdf.

30. 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;97(1):42-44.

31. Keller M, Freifeld CC, Brownstein JS. Automated vocabulary discovery for geo-parsing online epidemic intelligence. *BMC Bioinformatics*. 2009;10:385.

32. Brownstein JS, Freifeld CC. Evaluation of Internet-Based Informal Surveillance for Global Infectious Disease Intelligence. *International Journal of Infectious Diseases*. 2008;12(Supplement 1):e193-e194.

33. Brownstein JS, Freifeld CC, Reis BY, Mandl KD. Evaluation of online media reports for global infectious disease intelligence. *Advanced in disease surveillance 2007;4:3*. 2007. Available at: http://www.isdsjournal.org/articles/1935.pdf. Accessed November 22, 2010.

34. Chen H, Yan P, Zeng D, et al. HealthMap. In: *Infectious Disease Informatics*.Vol 21. Integrated Series in Information Systems. Springer US; 2010:183-186. Available at: http://dx.doi.org/10.1007/978-1-4419-1278-7_14.

35. Freifeld CC, Mandl KD, Reis BY, Brownstein JS. HealthMap: Global Infectious Disease Monitoring through Automated Classification and Visualization of Internet Media Reports. *Journal of the American Medical Informatics Association*. 2008;15(2):150-157.

36. Brownstein JS, Freifeld CC. HealthMap: the development of automated real-time internet surveillance for epidemic intelligence. *Euro Surveill*. 2007;12(11):E071129.5.

37. Nelson R. HealthMap: the future of infectious diseases surveillance? *The Lancet Infectious Diseases*. 2008;8(10):596.

38. Brownstein JS, Freifeld CC, Reis BY, Mandl KD. Surveillance Sans Frontières: Internet-based emerging infectious disease intelligence and the HealthMap project. *PLoS Med*. 2008;5(7):e151.

39. Kass-Hout TA, di Tada N. International System for Total Early Disease Detection (InSTEDD) Platform. 2009. Available at: http://www.isdsjournal.org/article/viewArticle/3308. Accessed November 19, 2010.

40. Yangarber R, Steinberger R, Best C, et al. Combining Information Retrieval and Information Extraction for Medical Intelligence. *Proceeding of Mining Massive Data Sets for Security, NATO Advanced Study Institute. Gazzada, Italy 2007*. 2007. Available at: http://langtech.jrc.it/mmdss2007/htdocs/Presentations/Docs/MMDSS_Yangarber_Steinberger_PUBLIC.pdf.

41. Rortais A, Belyaeva J, Gemo M, van der Goot E, Linge JP. MedISys: An early-warning system for the detection of (re-)emerging food- and feed-borne hazards. *Food Research International*. 2010;43(5):1553-1556.

42. STEINBERGER R, FUART F, van der GOOT E, et al. Text Mining from the Web for Medical Intelligence. 2008. Available at: http://langtech.jrc.it/Documents/2009_MMDSS_Medical-Intelligence.pdf. Accessed November 19, 2010.

43. Damianos L, Ponte J, Wohlever S, et al. MiTAP for Biosecurity: A Case Study. *AI Magazine*. 2002;23(4):13-29.

44. Hugh-Jones M. Global awareness of disease outbreaks: The experience of ProMED-mail. *Public Health Reports*. 2001;116(SUPPL. 2):27-31.

45. Woodall JP. Global surveillance of emerging diseases: the ProMED-mail perspective. *Cad Saude Publica*. 2001;17 Suppl:147-154.

46. Woodall J. Official versus unofficial outbreak reporting through the Internet. *Int J Med Inform*. 1997;47(1-2):31-34.

47. Morse SS, Rosenberg BH, Woodall J. ProMED global monitoring of emerging diseases: Design for a demonstration progam. *Health Policy*. 1996;38(3):135-153.

48. Madoff LC. ProMED-mail: an early warning system for emerging diseases. *Clin. Infect. Dis*. 2004;39(2):227-232.

49. Woodall J. Stalking the next epidemic: ProMED tracks emerging diseases. *Public Health Reports*. 1997;112(1):78-82.

50. Madoff LC, Woodall JP. The internet and the global monitoring of emerging diseases: lessons from the first 10 years of ProMED-mail. *Arch. Med. Res*. 2005;36(6):724-730.

51. Zeldenrust ME, Rahamat-Langendoen JC, Postma MJ, van Vliet JA. The value of ProMED-mail for the Early Warning Committee in the Netherlands: more specific approach recommended. *Euro Surveill*. 2008;13(6). Available at: http://www.ncbi.nlm.nih.gov/pubmed/18445424. Accessed November 18, 2010.

52. Grishman R, Huttunen S, Yangarber R. Information extraction for enhanced access to disease outbreak reports. *J Biomed Inform*. 2002;35(4):236-246.

53. Anon. Hartley D. one year later: implementing the bio-surveillance requirement of the 9/11 Act. Statement by Dr David Hartly, Georgetown University Medical Center. Hearing before the Subcomm. On Emerging Threats, Cybersecurity, and Science

and Technology of the house Comm. On Homeland Security, 110th Cong., 2nd Sess. (July 16, 2008). Available at: http://www.fas.org/irp/congress/2008_hr/biosurv.pdf. Accessed July 14, 2011.

54. Conway M, Doan S, Kawazoe A, Collier N. Classifying disease outbreak reports using n-grams and semantic features. *International Journal of Medical Informatics*. 2009;78(12):e47-e58.

55. Collier N. What's unusual in online disease outbreak news? *J Biomed Semantics*. 2010;1(1):2.

56. Anon. BioCaster Portal - About BioCaster. Available at: http://born.nii.ac.jp/?page=about. Accessed July 15, 2011.

57. Collier N, Kawazoe A, Jin L, et al. A Multilingual Ontology for Infectious Disease Surveillance: Rationale, Design and Challenges. *Language Resources and Evaluation*. 2006;40(3/4):405-413.

58. Coulombier D. Epidemic intelligence in the European Union: strengthening the ties. *Euro surveillance : bulletin européen sur les maladies transmissibles = European communicable disease bulletin*. 2008;13(6). Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-44349111160&partnerID=40&md5=6bd9c30ad071ef57a78943477359c7bc. Accessed May 4, 2011.

59. Anon. Independent Evaluation of the GOARN Network. Available at: http://www.hlsp.org/Home/Projects/GOARNEvaluation.aspx. Accessed September 11, 2011.

60. Grein TW, Kamara K-BO, Rodier G, et al. Rumors of disease in the global village: Outbreak verification. *Emerging Infectious Diseases*. 2000;6(2):97-102.

61. Anon. InSTEDD | Innovative Support to Emergencies Diseases and Disasters. Available at: http://instedd.org/. Accessed July 25, 2011.

62. Anon. Europe Media Monitor - EMM. Available at: http://emm.newsbrief.eu/overview.html. Accessed August 13, 2011.

63. Anon. Combining Information Retrieval and Information Extraction for Medical Intelligence. Available at: http://langtech.jrc.it/mmdss2007/htdocs/Presentations/Docs/MMDSS_Yangarber_Steinberger_PUBLIC.pdf. Accessed August 14, 2011.

64. Anon. PULS Project: University of Helsinki. Available at: http://puls.cs.helsinki.fi/medical/. Accessed August 13, 2011.

65. Anon. JRC-PULS. Available at: http://sysdb.cs.helsinki.fi/tomcat/tkt_plus/jrc/. Accessed August 13, 2011.

66. Mani I, Bloedorn E. Summarizing Similarities and Differences Among Related Documents. *Information Retrieval*. 1999;1(1/2):35-67.

67. Cowen P, Garland T, Hugh-Jones ME, et al. Evaluation of ProMED-mail as an electronic early warning system for emerging animal diseases: 1996 to 2004. *Journal of the American Veterinary Medical Association*. 2006;229(7):1090-1099.

68. Anon. About ProMED-mail. Available at: http://www.promedmail.org/. Accessed August 23, 2011.

69. Chan EH, Brewer TF, Madoff LC, et al. Global capacity for emerging infectious disease detection. *Proceedings of the National Academy of Sciences*. 2010;107(50):21701 -21706.

70. Anon. Global distribution of diseases: TB, HIV/Aids, cholera and malaria - MyFundi. Available at: http://myfundi.co.za/e/Global_distribution_of_diseases:_TB,_HIV/Aids,_cholera_and_malaria. Accessed September 17, 2011.

71. European Centre for Disease Prevention and Control (ECDC) - Health Comunication Unit - Eurosurveillance editorial team. What is epidemic intelligence, and how is it being improved in Europe? 2006. Available at: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=2892. Accessed May 16, 2011.

72. 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.