Global Public Health Intelligence Network (GPHIN)

Abla Mawudeku
Chief, Global Public Health Intelligence Network (GPHIN)
100 Colonnade Road
A.L. 6201A
Ottawa, ON
K1A 0K9
Canada
abla_mawudeku@phac-aspc.gc.ca

Michael Blench
GPHIN Technical Advisor and Project Coordinator
100 Colonnade Road
A.L. 6201A
Ottawa, ON
K1A 0K9
Canada
michael_blench@phac-aspc.gc.ca

Abstract

Accurate and timely information on global public health issues is key to being able to quickly assess and respond to emerging health risks around the world. The Public Health Agency of Canada has developed the Global Public Health Intelligence Network (GPHIN). Information from GPHIN is provided to the WHO, international governments and non-governmental organizations who can then quickly react to public health incidents.

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. This unique multilingual system gathers and disseminates relevant information on disease outbreaks and other public health events by monitoring global media sources such as news wires and web sites. This monitoring is done in eight languages with machine translation being used to translate non-English articles into English and English articles into the other languages. The information is filtered for relevancy by an automated process which is then complemented by human analysis. The output is categorized and made accessible to users. Notifications about public health events that may have serious public health consequences are immediately forwarded to users.

GPHIN employs a "best-of-breed" approach when it comes to the selection of the machine translation 'engines'. This philosophy ensures that the quality of the machine translation is the best available for whatever language pair selected. It also imposes some unique integration and operational problems.

GPHIN has a broad scope. It tracks events such as disease outbreaks, infectious diseases, contaminated food and water, bio-terrorism and exposure to chemicals, natural disasters, and issues related to the safety of products, drugs and medical devices.

GPHIN is managed by Health Canada's Centre for Emergency Preparedness and Response (CEPR), which was created in July 2000 to serve as Canada's central coordinating point for public health security. It is considered a centre of expertise in the area of civic emergencies including natural disasters and malicious acts with health repercussions. CEPR offers a number of practical supports to municipalities, provinces and territories, and other partners involved in first response and public health security. This is achieved through its network of public health, emergency health services, and emergency social services contacts.

1 Introduction

The globalization of the world economies and societies has presented the world order with new challenges. Today's world is more interconnected on issues of health, economics and trade. One of the effects of globalization has been increased international travel and trade. This however, has increased the microbial threat worldwide making nations vulnerable to public health threats. Pathogens can now be easily transported across nations via vectors such as humans or cargo, facilitating their ability to invade new territory and adapt to new environments and hosts. As a result,

nations need to consider emerging public health threats worldwide when determining what potential microbial threats may affect the health of their nation.

In order to keep abreast of potential public health threats worldwide and undertake prompt prevention and control activities, global surveillance has to be timely and comprehensive in the gathering and dissemination of information to public health officials. This is possible in most developed countries however, in countries where the public health infrastructures are rudimentary, deteriorating or non-existent, reporting of many public health threats is considerably less than adequate. Furthermore, the reluctance of some countries or authorities to report potential threats due to the negative impact on trade and tourism, or to gain a tactical advantage, has also resulted in limited exchange of information between authorities.

A new means to address these challenges is therefore needed in order to strengthen the global public health surveillance system. One approach is to take advantage of the advancements in communication technologies. The internet revolution has made the world more connected and has enhanced the practice of public health surveillance. A distributed system of coordinated, timely and useful multi-source public health information can now be more readily developed. An example of such a system taking advantage of today's advancements in communication technologies is the Global Public Health Intelligence Network (GPHIN).

2 Background

The prototype GPHIN system was developed in 1997 in partnership with the World Health Organization (WHO). The objective was to determine if the use of the internet to continuously monitor and gather information about possible disease outbreaks worldwide and be able to alert international bodies of such events in a timely fashion was feasible and beneficial. The system scanned global media sources worldwide in English and French. The sources included websites, news wires, local and national newspapers.

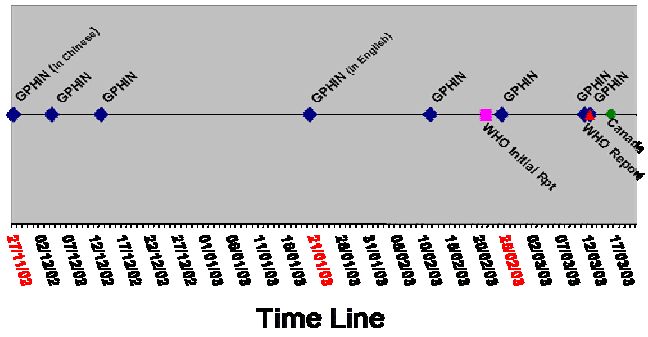
2.1 Evaluation and Utility

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[1]. The outbreaks reported were occurring in 132 countries around the world demonstrating GPHIN's capacity to monitor worldwide despite the limitation of monitoring media sources only in English and French.

During the SARS outbreak in 2002, the GPHIN prototype demonstrated its effectiveness as an early-warning system. It was able to gather information about an unusual outbreak occurring in Guangdong Province, Mainland China, as early as November 27, 2002. At that time, GPHIN had begun monitoring media sources in the six WHO official languages – Arabic, Chinese, English, French, Russian and Spanish. However, the GPHIN system could not accommodate the presentation of information in the other languages besides English and French. GPHIN Analysts therefore translated only the titles of selected non-English articles. The translated titles and the articles in their original language were then manually forwarded to users via email.

Figure 1 demonstrates the importance of monitoring in various languages and types of news content. The first reports of the unusual outbreak were in Chinese and it was only in January 21, 2003 that there was a first report in English. Also, this was a financial report about a pharmaceutical company's increased sales of antivirals in the Guangdong Province, Mainland China. The report attributed the increased sales to the unusual outbreak occurring in that region. The WHO released its first report on February 25, 2003 informing the public of an atypical pneumonia outbreak.

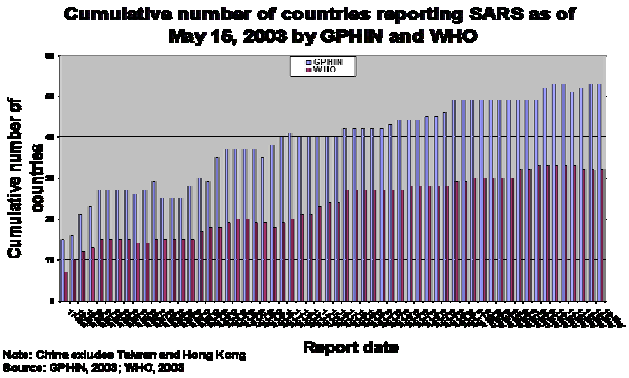
Figure 1



GPHIN was able to continuously monitor and provide information about the number of suspected and probable SARS cases reported worldwide on a near-real time basis. GPHIN information about cases was approximately two to three days ahead

of the official WHO report of cases worldwide (see Figure 2). Although the GPHIN information was unverified and the magnitude of cases reported was different, the epidemic curve was similar to that of the WHO.

Figure 2



GPHIN also provided information that enabled public health officials to:

- Identify control and preventative measures that were considered and implemented worldwide;
- Monitor concerns of the general public; and
- Observe situational politics related to the SARS outbreak.

3 GPHIN: A Multilingual Early Warning System for Global Public health Surveillance

Following the SARS outbreak, development of the new GPHIN multilingual system commenced in order to replace the prototype which was not robust enough to continue supporting global public health surveillance activities. The new system was launched on November 17, 2004 at the United Nations. It is a secure Internet-based "early warning" system that gathers preliminary reports of public health significance on a near "real-time" basis, 24 hours a day, 7 days a week. This unique multilingual system gathers and disseminates relevant information on disease outbreaks and other public health events by monitoring global media sources such as news wires and web sites. The information is filtered for relevancy by an automated process which is then complemented by human analysis. The output is categorized and made accessible to users. Notifications about public health events that may have serious public health consequences are immediately forwarded to users via email.

GPHIN covers a broad scope of public health issues. It tracks events such as infectious disease outbreaks, contaminated food and water, bio-terrorism, chemical and radioactive incidents, natural disasters and issues related to the safety of products, drugs and medical devices.

4 Processing Information

4.1 Automated Process

The GPHIN software application 'pulls' relevant articles every 15 minutes from newsfeed aggregators (Al Bawaba and Factiva) based on established search syntaxes that are updated regularly as necessary. The articles are further filtered and categorized into one or more of GPHIN's taxonomy categories which cover the following:

- Animal diseases
- Human diseases
- Plant diseases
- Biologics
- Natural disasters
- Chemical incidents
- Radioactive incidents
- Unsafe products

Each article is assigned a 'relevancy score' which is derived from a proprietary algorithm utilizing the values attributed to the keywords and terms within the taxonomy or taxonomies it has been assigned to.

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 is below an established threshold are automatically 'trashed' as being not relevant according to GPHIN criteria.

4.2 Human Analysis Process

While the sophisticated GPHIN computer system is essential for the management of information about health threats worldwide, it is the linguistic, interpretive and analytical expertise of the GPHIN analysts that makes the system a success.

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. Subsequent to reviewing an article, the analyst decides 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 indepth analysis including horizontal analysis linking events in different regions, identifying and reporting on trends and assessing the health risks to the populations around the world.

4.3 Machine Translation

English articles are machine translated into Arabic, Chinese (Simplified), Farsi, French, Russian and Spanish. Non English articles, which include Chinese (Traditional), are machine translated into English.

Currently, the most advanced machine translation softwares are very imperfect. As such, the 'Gists', or results of the machine translation, are presented to the appropriate GPHIN analysts to improve their comprehensibility. The analyst is not mandated to do a perfect translation, but to provide the essence of the translated article.

The machine translation 'engines' use dictionaries which are constantly being refined by expert linguists and the GPHIN analysts. Over time, the quality of the machine translations will improve.

GPHIN employs a "best-of-breed" approach when it comes to the selection of the machine translation 'engines'. We are in constant touch with industry to ensure that we incorporate the very best of what is available.

The "best-of-breed" approach presents increased integration issues as well as augmented operational issues especially when dealing with as many 'engines' as GPHIN does. These issues include items such as instability, crashes, unpredictable performance, poor documentation, awkward API's (Application Program Interface), lack of standards across products and the inevitable 'bugs'.

GPHIN has devised a software module that greatly improves some of these issues. This module called GIST-IN-TIME, normalizes most of the API's, detects when an 'engine' has crashed and re-boots it and overcomes some of the incompatibilities between / amongst some of the 'engines' when they are co-resident on the same server.

5 Access to Information

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.

The query function permits any combination or permutation of exact character matches (Boolean logic) with metadata matches from the GPHIN taxonomy which contains all the terms / keyword in all languages, all synonyms and all spelling variation of all terms / keywords. This taxonomy is constantly being revised as new terms come into vogue.

6 GPHIN Platform

The GPHIN platform can be customized to meet the requirements of a specific program (see Figure 3 on next page). Specific adaptation of the taxonomy and search syntaxes in the languages of choice can be done. The sources of information can be expanded to accommodate a broad spectrum of data types which include for example, classified/private information, websites and listserves.

Although GPHIN is currently operating as a stand alone application, a high degree of interoperability can be achieved due to the use of standardized data formats – XML, HTML, etc.

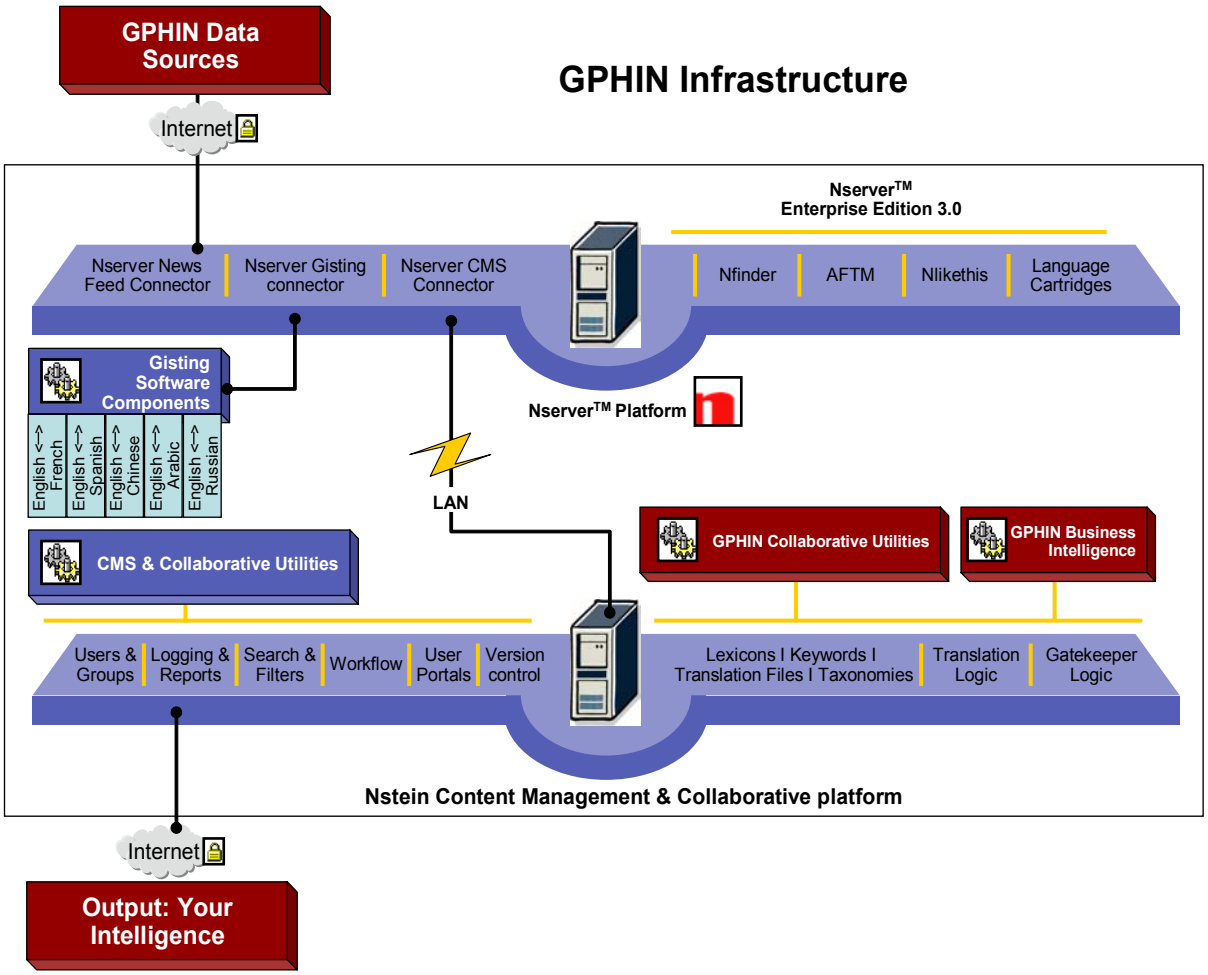
7 Conclusion

The Global Public Health Intelligence Network – both the technology and the analytical team that drives the technology – has proven to be an effective tool in the detection and management of public health threats.

References

1. Heymann DL, Rodier GR, et al. Hot Spots in a wired world: WHO surveillance of emerging and re-emerging infectious diseases, *The Lancet*; 1:345-353.

Figure 3


GPHIN Data Sources
GPHIN Infrastructure
Internet
Nserver™ Enterprise Edition 3.0
Nserver News Feed Connector
Nserver Gisting connector
Nserver CMS Connector
Nfinder
AFTM
Nlikethis
Language Cartridges
Gisting Software Components
English <-> French
English <-> Spanish
English <-> Chinese
English <-> Arabic
English <-> Russian
Nserver™ Platform
LAN
CMS & Collaborative Utilities
GPHIN Collaborative Utilities
GPHIN Business Intelligence
Users & Groups
Logging & Reports
Search & Filters
Workflow
User Portals
Version control
Lexicons I Keywords I Translation Files I Taxonomies
Translation Logic
Gatekeeper Logic
Nstein Content Management & Collaborative platform
Internet
Output: Your Intelligence