

KIWI: A technology for public health event monitoring and early warning signal detection

Shamir Mukhi¹, Ellie Andres¹, Bryan Demianyk¹, Ben Gammon¹, Harold Kloeze²

1. Canadian Network for Public Health Intelligence, National Microbiology Laboratory, Winnipeg, Manitoba
2. Canadian Food Inspection Agency, Owen Sound, Ontario

Abstract

Objectives: To introduce the Canadian Network for Public Health Intelligence's new Knowledge Integration using Web-based Intelligence (KIWI) technology, and to perform preliminary evaluation of the KIWI technology using a case study. The purpose of this new technology is to support surveillance activities by monitoring unstructured data sources for the early detection and awareness of potential public health threats.

Methods: A prototype of the KIWI technology, adapted for zoonotic and emerging diseases, was piloted by end-users with expertise in the field of public health and zoonotic/emerging disease surveillance. The technology was assessed using variables such as geographic coverage, user participation, and others; categorized by high-level attributes from evaluation guidelines for internet based surveillance systems. Special attention was given to the evaluation of the system's automated sense-making algorithm, which used variables such as sensitivity, specificity, and predictive values. Event-based surveillance evaluation was not applied to its full capacity as such an evaluation is beyond the scope of this paper.

Results: KIWI was piloted with user participation = 85.0% and geographic coverage within monitored sources = 83.9% of countries. The pilots, which focused on zoonotic and emerging diseases, lasted a combined total of 65 days and resulted in the collection of 3243 individual information pieces (IIP) and 2 community reported events (CRE) for processing. Ten sources were monitored during the second phase of the pilot, which resulted in 545 anticipatory intelligence signals (AIS). KIWI's automated sense-making algorithm (SMA) had sensitivity = 63.9% (95% CI: 60.2-67.5%), specificity = 88.6% (95% CI: 87.3-89.8%), positive predictive value = 59.8% (95% CI: 56.1-63.4%), and negative predictive value = 90.3% (95% CI: 89.0-91.4%).

Discussion: Literature suggests the need for internet based monitoring and surveillance systems that are customizable, integrated into collaborative networks of public health professionals, and incorporated into national surveillance activities. Results show that the KIWI technology is well poised to address some of the suggested challenges. A limitation of this study is that sample size for pilot participation was small for capturing overall readiness of integrating KIWI into regular surveillance activities.

Conclusions: KIWI is a customizable technology developed within an already thriving collaborative platform used by public health professionals, and performs well as a tool for discipline-specific event monitoring and early warning signal detection.

Keywords: Digital disease detection, event-based surveillance, internet-based surveillance, epidemiology, text mining, biosurveillance.

Correspondence: shamir.nizar.mukhi@phac-aspc.gc.ca

DOI: 10.5210/ojphi.v8i3.6937

Copyright ©2016 the author(s)

This is an Open Access article. Authors own copyright of their articles appearing in the Online Journal of Public Health Informatics. Readers may copy articles without permission of the copyright owner(s), as long as the author and OJPHI are acknowledged in the copy and the copy is used for educational, not-for-profit purposes.

Background

Internet Biosurveillance

Internet biosurveillance emerged in the mid-1990s and has matured into a globally recognized technique for providing early warning of, and situational awareness for, public health threats [1-2]. This web-based approach utilizes unstructured real-time or near real-time data to support and complement traditional indicator-based surveillance. Many internet biosurveillance systems have been developed with examples of well-established and analyzed systems including Argus, BioCaster, EpiSPIDER, GPHIN, HealthMap, MedISys, and ProMED-mail [3-8].

Since internet biosurveillance, or non-traditional event-based surveillance, is distinct from traditional indicator-based surveillance, it is recognized that applying the Centers for Disease Control and Prevention's "Updated Guidelines for Evaluating Public Health Surveillance Systems" [9] to the evaluation of internet biosurveillance systems is not suitable [10]. Therefore, criteria specific to event-based surveillance systems were developed during a workshop in 2010, which was held at the Pacific Northwest National Laboratory in Richland, Washington. These criteria include the following attribute families: Event (i.e. description of event including source, causative agent, and detection mode), Readiness (i.e. system validation and stakeholder willingness to use the system), Operational Aspects (e.g. administration/maintenance requirements, and system redundancy and ability to accommodate various levels of data), Geographic Coverage, Population Coverage, Input Data (e.g. accessibility, quality, quantity, and utility of input data), Output (e.g. accessibility, quality, quantity, and utility of output data), and Cost (i.e. funding sustainability, and research, evaluation, and operational expenses).

Six of these eight attribute families were used to guide categorization of KIWI evaluation results based on a specific case study. Cost and Operational Aspects were not direct results of this evaluation; however, operational aspects may be identified throughout the introduction of KIWI.

Internet biosurveillance systems have been compared to one another and evaluated both qualitatively and quantitatively throughout literature and system challenges are readily discussed [3-8]. In reference to event-based biosurveillance systems, Keller and colleagues suggested in 2009 that the future development of event-based systems should focus on establishing

collaborative networks of public health practitioners for the verification and dissemination of early warning signals [3]. In 2010, Hartley and colleagues suggested that “prominent challenges [with event-based systems] include interoperability, interface customizability, scalability, and event traceability” [4]. In addition, it was suggested that biosurveillance capability can be expanded with the use of emerging media such as social networking sites and that the similarities and differences between event-based systems indicate that a more powerful resource can be created by combining them. The idea of combining existing internet biosurveillance systems to create a stronger platform is echoed throughout the literature [4, 6-7]. In 2013, Hartley and colleagues suggested the use of interactive functions for users such as scoring options and comment fields [1]. In 2014, a systematic review by Velasco and colleagues assessed 13 event-based surveillance systems from Canada, the European Union, Japan, and the United States and identified that no system had been incorporated into a national surveillance program [8].

In summary, literature suggests the need for:

- Enhancing technologies including the ability to customize the system’s user interface, trace events from beginning to end, adjust to various volumes of data input and coverage (scalability), and be functional despite jurisdictional boundaries (interoperability);
- Establishing collaborative networks of public health professionals for the verification and dissemination of early warning signals and addition of interactive functions;
- Combining existing event-based systems; and
- Integrating event-based systems into national surveillance initiatives.

KIWI is uniquely designed within an existing national surveillance platform and is well poised to address the additional challenges proposed in literature regarding internet based monitoring and surveillance systems.

Canadian Network for Public Health Intelligence

The Canadian Network for Public Health Intelligence (CNPHI) is a Public Health Agency of Canada (PHAC) initiative developed and managed by the National Microbiology Laboratory (NML) [11]. CNPHI is a secure and comprehensive framework of applications and resources designed to enable multi-jurisdictional surveillance, response, and collaboration in the field of public health; CNPHI supports initiatives ranging from zoonotic disease detection to drinking water advisories, nosocomial infections, food borne illnesses and many others. The platform was established in 2003 and is built upon six focus areas: Knowledge Management, Surveillance, Alerting, Collaboration, Event Management, and Laboratory Systems.

To complement and support surveillance activities currently performed using the CNPHI platform, an event-based monitoring application called “Knowledge Integration using Web-based Intelligence” (KIWI) has been developed within the Knowledge Management focus area.

The Knowledge Integration using Web-based Intelligence Technology

The purpose of CNPHI’s KIWI technology is to support surveillance activities by monitoring events using unstructured data sources for the early detection and awareness of potential public health threats. KIWI is designed to collect information from various internet sources and process intelligence using an automated sense-making algorithm (SMA). Individual information pieces (IIPs) are raw individual information items from RSS feeds. Once processed by KIWI’s automated SMA, IIPs either remain IIPs or become potential early warning signals, also referred to as anticipatory intelligence signals (AIS). AISs are presented to a community of CNPHI users for manual relevancy rating. Highly rated AISs become early warning signals (EWS), which are then disseminated to the user community. Users are also encouraged to record community reported events (CRE), which are entered into the KIWI system as additional unprocessed AISs for community rating. Figure 1 is a schematic of the information flow that underpins KIWI.

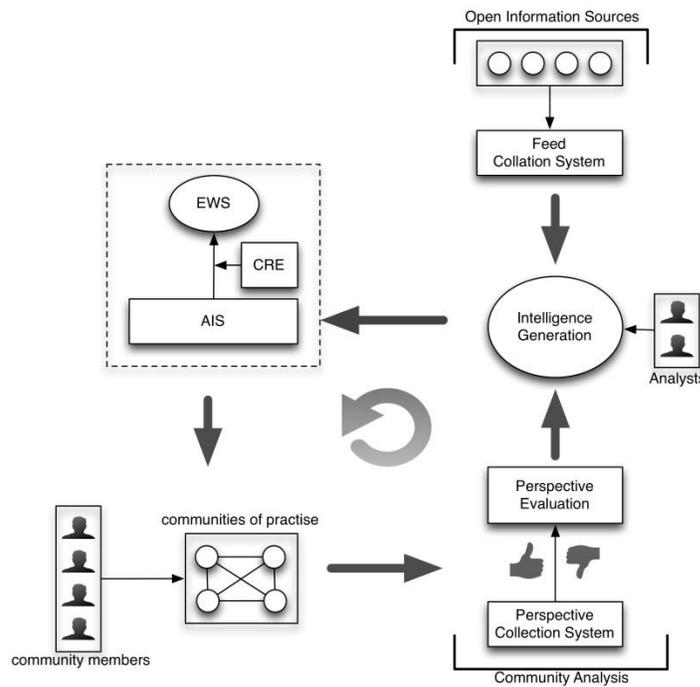


Figure 1. KIWI Information Flow.

KIWI takes advantage of both automated and manual processing to reduce the amount of time and resources required for reviewing IIPs and to ensure EWS relevance through manual validation. The KIWI interface presents IIPs and AISs/EWSs, or signals, in both map and listing formats with many search and filtering options. The map format allows users to view signals by geography, while the listing format allows users to view details including, but not limited to, title, description, source, and full-text link. Each signal is accompanied by supporting information via health condition specific links to tools such as Google Trends and the Global Infectious Diseases and Epidemiology Network (GIDEON). Each signal is equipped with forums for community interaction and tools for following related signals over time. In addition, if users

want to remain updated on specific signals, they can “watch” the signal and receive email notifications.

Since CNPHI users are multi-disciplinary and multi-jurisdictional, KIWI can be customized to meet the requirements of various organizations and collaborations. Figure 2 displays an overview of KIWI’s high-level components including data collection, processing, analysis, and dissemination. Details such as specific data sources and dictionaries (keywords and weights) vary by KIWI program (ex: Zoonotic).

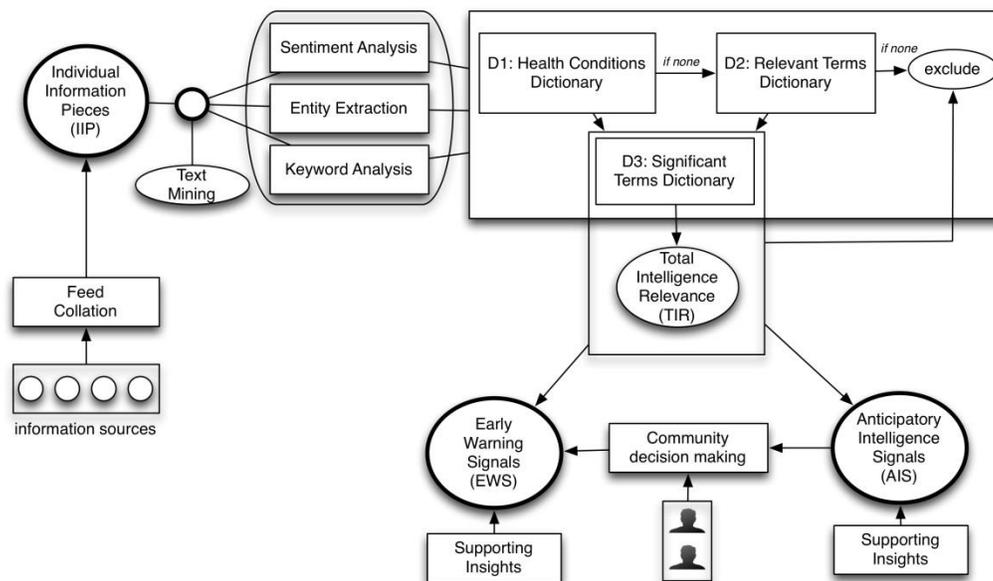


Figure 2. An overview of KIWI’s technology

Collection and Storage

RSS feeds are used to collect information in the form of IIPs from internet-based sources such as Eurosurveillance, MedISys, ProMED-mail, and others. IIPs are indexed and made available directly on the KIWI platform by use of customized searches.

Data Processing

Each IIP is fed through text mining software (Alchemy) to extract keywords, entities (e.g. geography), and keyword characteristics such as relevance and sentiment. Entities are used to map IIPs when geography is provided and gives details for tabulation. Identified keywords are accompanied by values indicating percent relevancy (R) and sentiment score (S). Each keyword is then searched in pre-assembled KIWI dictionaries where matched keywords are given additional weight (W), which can be either a positive or negative value. The following formula is used to calculate Total Intelligence Relevance (TIR): $\sum (R - S) * W$, where TIR is the sum of

keyword relevance minus sentiment multiplied by weight for each keyword. TIR is used to determine whether an IIP becomes an AIS or EWS.

Within the KIWI technology, duplicate IIPs may occur from the same source or from different sources. The current implementation does not automatically remove duplicates, however, the system flags similar IIPs that have occurred within the last 4 weeks. Further work in this area could be beneficial.

Data Analysis

AISs are manually rated by users on a scale of one to five, with one being not-relevant and five being extremely relevant (Figure 3). IIPs rated higher than not-relevant are considered valid AISs, and those rated greater or equal to relevant are considered valid EWSs. Note that valid AISs become EWSs automatically if 60% of users rate greater than or equal to relevant with a minimum of 10 raters, otherwise manual validation takes place by analysts. Community rating may be used as a gold standard for calculating sensitivity, specificity and predictive values for the sense-making algorithm.

Dissemination

KIWI includes an interactive interface to view signals and the technology is capable of creating auto-generated reports summarizing AISs and validated EWSs. Generated reports may be disseminated to appropriate communities via associated CNPHI collaboration centre, which include options for managing documents and notifying select workgroups via email. Auto-generated reports may include the following signal-related information: title, description, date posted, program, source, signal type (AIS/EWS/CRE), detection method, and average rating.

Case Study

Though KIWI is designed to accommodate various programs, or topics/disciplines, the described study focuses solely on the Zoonotic program. The Zoonotic program was customized in collaboration with the Centre for Emerging and Zoonotic Disease Integrated Intelligence and Response (CEZD-IIR). For the purpose of this paper, this customization of KIWI will be referred to as KIWI-Zoonotic.

KIWI-Zoonotic was piloted in two phases between June and November of 2015, with each phase lasting approximately one month in duration. The purpose of the first phase was to familiarize pilot participants with the technology, and verify its functionality and usability, while the second phase aimed to measure KIWI performance. Pilot participants (Phase I: n = 20, Phase II: n = 37) were real end-users of zoonotic/emerging disease intelligence including veterinarians, epidemiologists, scientists, analysts, and others from local, provincial and federal institutions and agencies across Canada.

Preparation

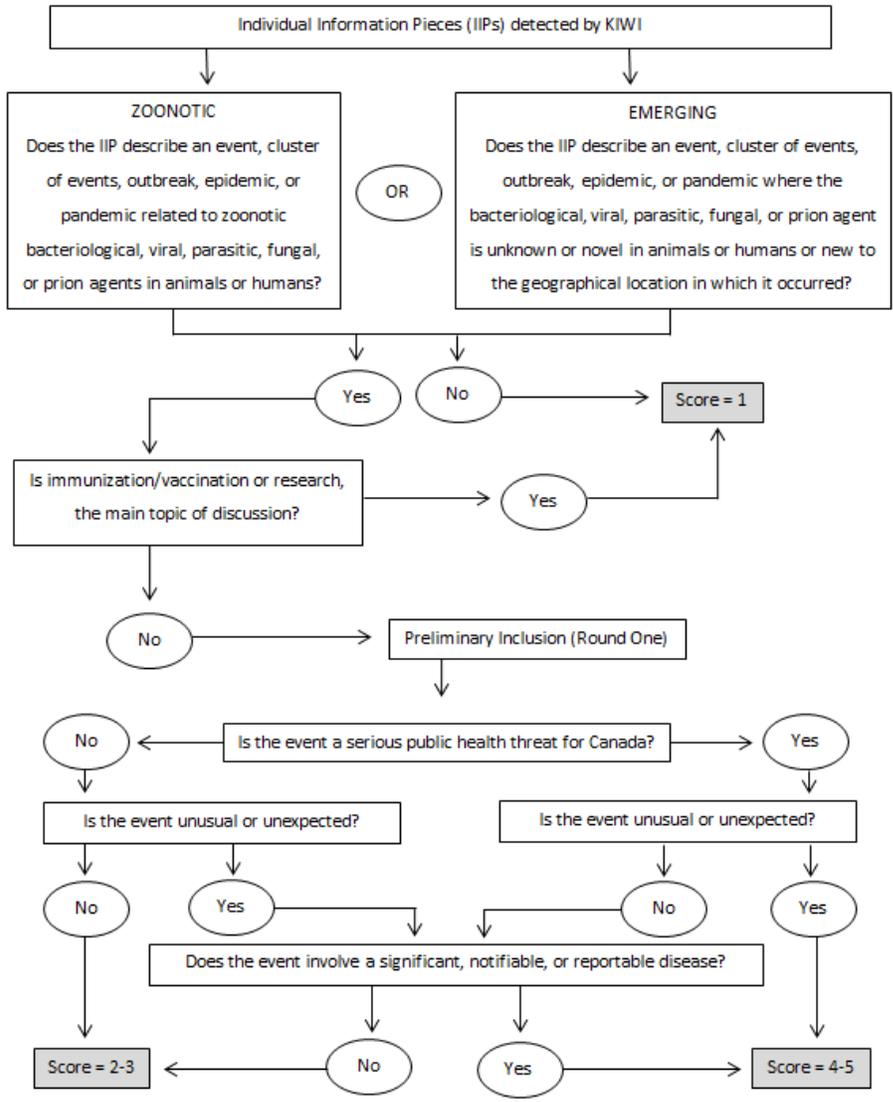
In preparation of the KIWI-Zoonotic pilot, data sources, dictionaries, and relevancy rating decision making criteria were configured to meet CEZD-IIR requirements. Data sources were identified and a standardized tool for prioritizing information sources was applied to refine selected sources[12]. KIWI-Zoonotic utilizes the following internet-based information sources: EurekAlert, Eurosurveillance, IBIS, MedISys, Outbreak News, Pig Process, ProMED-mail, Science Daily, and The Poultry Site.

The KIWI technology requires three dictionaries: Health Conditions, Relevant Terms, and Significant Terms.

- Health Conditions is a list of known health conditions and diseases of interest with assigned weights reflecting relevance. For the Zoonotic program, this dictionary contains known zoonotic and known emerging diseases. Keyword weights were determined by sorting diseases by their presence or absence on various notifiable disease lists (animal/human and provincial/national/international). These weights were then adjusted based on relevance to the community of users via feedback from program team leads.
- Relevant Terms is a list of terms used as a proxy for unknown health conditions to identify potential signals of interest with emerging capacity. For the Zoonotic program, this dictionary contains disease agents such as viruses, bacteria and others. Keywords were assigned neutral weight as there is no hierarchy in relevance of disease agents.
- Significant Terms is a list of terms used to define signal importance, such as, outbreaks, unknown diseases, new diseases, et cetera. For the Zoonotic program, Keywords were grouped by the following categories: exclusion terms, epidemiological terms, and novel terms. Exclusion terms were given negative weight; epidemiological terms were given positive weight based on levels of keyword severity (for example, case versus outbreak versus pandemic); and novel terms were assigned positive weight without hierarchical variation.

For the purposes of the pilot, these dictionaries were configured for the detection of known zoonotic and animal diseases as well as significant and relevant terms utilized for the detection of emerging diseases.

A decision making tool was developed to aid KIWI-Zoonotic users in rating the relevancy of AISs. The tool was based on the International Health Regulations of 2005 [13] and adjusted for the purpose of zoonotic and emerging diseases (Figure 3).



1 – Not Relevant, 2 – Some Relevance, 3 – Relevant, 4 – Very Relevant, 5 – Extremely Relevant

Figure 3. Relevancy Assessment Tool for Zoonotic and Emerging Diseases

Evaluation Methods

Six of eight attribute families described earlier were used to categorize variables assessed during the KIWI-Zoonotic evaluation: events, geographic coverage, population coverage, readiness, data input, and data output.

Events were represented by the number of potential early warning signals detected during the two pilot phases. The number of IIPs indicates the pool of possible events while the number of potential early warning signals indicates the number of potentially relevant events. These variables were measured over both pilot periods.

Geographic coverage was calculated by identifying the number of countries referred to in detected AISs and dividing that number by the total number of countries. The numerator was determined by viewing KIWI-Zoonotic's map of potential early warning signals, and the denominator of 193 countries was based on members of the United Nations. This variable was measured over a one year time period (April 2015-2016).

Population coverage is a qualitative variable describing the population of interest including two population types: health conditions/diseases and species affected. This variable is not time dependent.

Readiness was represented by user participation in the KIWI-Zoonotic pilot. Descriptive statistics were used to identify the proportion of participants who accessed KIWI-Zoonotic, rated signals, commented on AISs, conducted searches for IIPs, and entered CREs. The proportion of participants who rated signals was measured over both pilot periods, while remaining variables were measured during Phase I because the purpose of Phase I and II differed.

Data Input was represented by source performance, which was assessed by calculating the number of AISs produced per source and plotting it against the proportion of AISs identified as relevant per source. Source performance was measured during Phase II of the pilot because sources monitored were modified based on outcomes of Phase I. Phase II data provided the most recent information.

Data Output was represented by assessing the automated SMA. KIWI's automated SMA for the zoonotic pilot was analyzed by calculating its sensitivity, specificity, and predictive values. Average community relevancy rating was used as the gold standard in these calculations.

- IIPs detected by the automated SMA (Automatic) that were rated "Not Relevant" were treated as false positives, and those rated higher than "Not Relevant" were treated as true positives.
- IIPs not detected by the automated SMA (Manual) that were rated "Not Relevant" were treated as true negatives, and those rated higher than "Not Relevant" were treated as false negatives.

Analysts reviewed IIPs for missed potential early warning signals on a daily basis and entered them manually for community rating. Community reported events are directly input into the system as AISs without being processed by the automated SMA. Since CREs are unprocessed, they were excluded from this portion of the analysis. Variables were measured over both pilot periods.

Results

IIPs – Events, Geographic Coverage, and Population Coverage

The KIWI-Zoonotic pilot lasted a combined total of 65 days (Phase I = 36 days; Phase II = 29 days) and resulted in the collection of 3243 IIPs (Phase I = 1618 IIPs; Phase II = 1625) and 2 CREs (Phase I = 1; Phase II = 1).

The KIWI-Zoonotic system detected events on a global scale with total geographic coverage, within monitored sources, at 83.9% of countries.

Since KIWI-Zoonotic focuses on animal and zoonotic disease, both animal (wild, agricultural, and domestic) and human events were captured by the system. The most frequently occurring IIPs were those referring to Dengue Fever, Avian Influenza, Ebola, Chikungunya Virus, and Zika Virus events.

Readiness

The rate of user participation during Phase I of the KIWI-Zoonotic pilot was 85.0% (17/20). 76.5% (13/17) of participants who accessed the system rated AISs, 47.1% (8/17) commented on AISs, 35.3% (6/17) conducted customized searches for IIPs, and 5.9% (1/17) entered CREs. During Phase II, 77.8% (28/36) of participants rated AISs.

During the pilot period, the average number of IIPs collected on a daily basis was 50, and the average number of potential early warning signals identified on a daily basis was 16. With the use of KIWI's automated SMA and analysts, there is a 68% reduction in the number of signals that users would be required to view and rate on a daily basis.

Sources – Data Input

KIWI-Zoonotic configured ten sources for data input including the following: ProMED-mail, Outbreak News, MedISys, Science Daily, IBIS, EurekAlert, The Poultry Site, Eurosurveillance, Pig Progress, and CREs. Phase I of the pilot used six of these ten sources including ProMED-mail, MedISys, EurekAlert, Eurosurveillance, Science Daily, and CREs, while Phase II used all ten sources. During Phase II of the pilot, 545 AISs were collected with the highest proportions of AISs input from ProMED-mail (34.3%; n = 187), Outbreak News (26.6%; n = 145), and MedISys (13.8%; n = 75). ProMED-mail and Outbreak News also produced the highest proportions of relevant AISs with 44.0% (142/323) and 30.0% (97/323) respectively. The denominator of 323 represents the number of AISs rated greater than “not relevant” by users.

Figure 4 displays each source with its corresponding AIS frequency and proportion of relevant AISs produced.

SMA – Data Output

During the KIWI-Zoonotic pilot (Phases I & II), a total of 3243 IIPs were processed by the automated SMA. 1025 processed IIPs became AISs (Phase I = 481; Phase II = 544) and 2218 IIPs remained IIPs. Of the 1025 processed AISs, 70.8% (726) were detected through KIWI's automated SMA and 29.2% (299) were identified manually by analysts, see Figure 5.

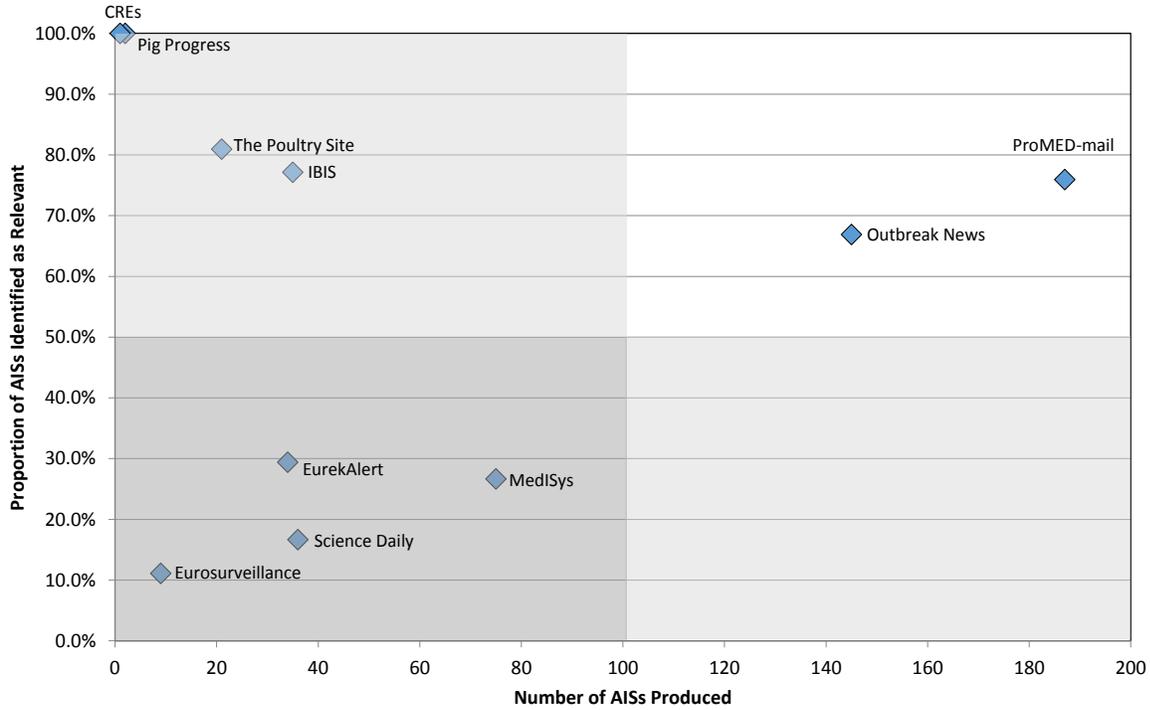


Figure 4. Proportion of Relevant Anticipatory Intelligence Signals by source and AIS frequency

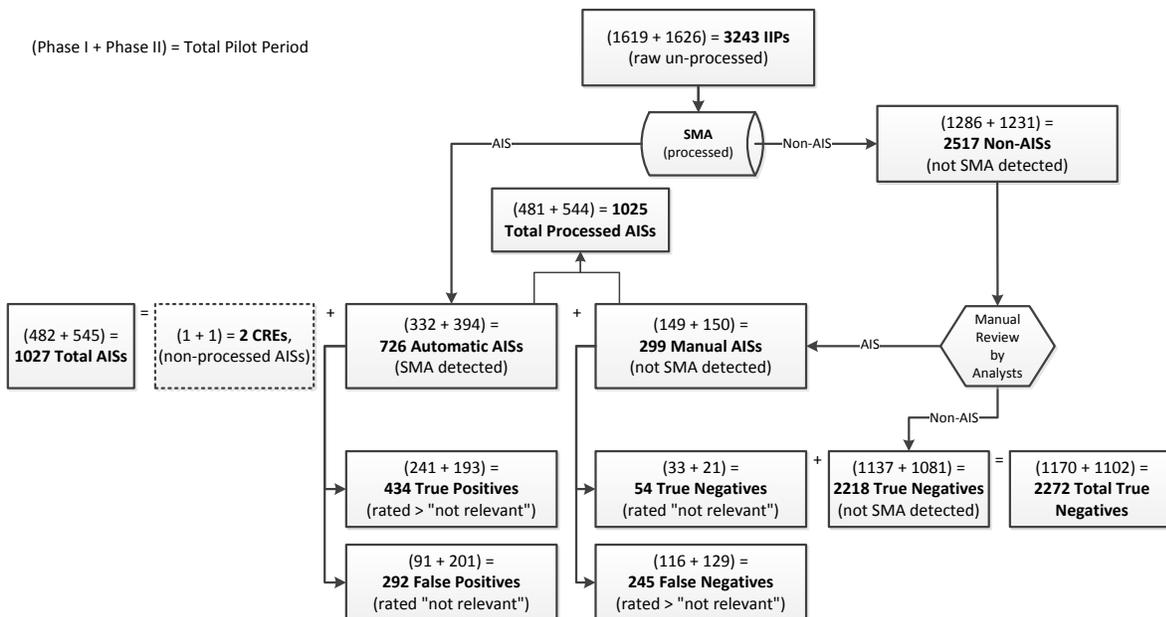


Figure 5. Breakdown of signals entering the KIWI during the pilot period

Of the 726 IIPs detected as AISs by the automated SMA, 434 were true positives and 292 were false positives (true positive rate = 59.8%; false positive rate = 40.2%). Of the 2517 IIPs not

detected as AISs by the automated SMA, 2272 were true negatives and 245 were false negatives (true negative rate = 90.3%; false negative rate = 9.7%), see Table 1.

Table 1. KIWI's Sense-Making Algorithm						
Zoonotic Pilot Phases I, II, & Total						
			AIS		Non-AIS	Total
Detected	Phase I		241		91	332
	Phase II		193		201	394
	Total		434		292	726
Not Detected	Phase I		116		1170	1286
	Phase II		129		1102	1231
	Total		245		2272	2517
Total	Phase I		357		1261	1618
	Phase II		322		1303	1625
	Total		679		2564	3243

The prevalence of AISs was 20.9% (679/3243). The probability that an AIS will be detected by the automated SMA is 63.9% (sensitivity; 434/679), and that a non-AIS will not be detected by the automated SMA is 88.6% (specificity; 2272/2564). The probability that a detected AIS will be a true positive is 59.8% (positive predictive value; 434/726), and that a non-detected IIP will be a true negative is 90.3% (negative predictive value; 2272/2517), see Table 2.

Table 2. KIWI's Sense-Making Algorithm			
Zoonotic Pilot Total Diagnostics			
		Percent	95% Confidence Interval
Prevalence		20.9%	19.6-22.4%
Sensitivity		63.9%	60.2-67.5%
Specificity		88.6%	87.3-89.8%
+ Predictive Value		59.8%	56.1-63.4%
- Predictive Value		90.3%	89.0-91.4%

Of 323 true AISs (detected manually, automatically, or via CRE) during Phase II of the pilot (1 CRE plus 322 total AIS Phase II; Table 1), 32 (9.9%) met the threshold for automatically becoming an EWS and an additional 37 (11.5%) met the criteria for manual EWS validation.

Discussion

KIWI-Zoonotic has broad geographic coverage, and processes individual information pieces, or IIPs, on a daily basis. Broad geographic coverage as a proportion of countries is important for the monitoring of diseases on a global scale despite population size or land mass. Increasing KIWI's geographic coverage is limited by the geographic reach of sources being monitored and event occurrence.

The system's automated SMA increases system usability by significantly decreasing the number of IIPs rated manually by users, which has been shown to increase user willingness to participate in the rating process. Since approximately one quarter of users did not contribute towards rating of the KIWI signals, assessment of factors contributing to user willingness to participate in the rating process and the further refinement of KIWI's SMA to reduce the number of false positives deserves future attention.

All KIWI-Zoonotic sources provided relevant information with varying proportions of relevant AIS production (Figure 4). ProMED-mail and Outbreak News ranked highly in both the proportion of relevant AISs and the number of AISs produced. Though CREs, Pig Progress, The Poultry Site, and IBIS each produced a small number of signals, their proportions of relevant signals were high. The remaining sources, EurekAlert, MediSys, Science Daily, and Eurosurveillance, each provided high proportions of non-relevant signals, or false positives, per source. False positive signals are useful for refining keyword dictionaries and the SMA's overall ability to distinguish relevant signals.

A study by Barboza and colleagues evaluated seven event-based systems (Argus, BioCaster, GPHIN, HealthMap, MedISys, ProMED, and Puls) on the following characteristics: Usefulness, Simplicity, Flexibility, Representativity, Completeness, Sensibility, and Timeliness [6].

Researchers concluded that no system ranked highly on every characteristic and thus systems with different strengths should be combined to make a stronger system. KIWI-Zoonotic takes advantage of this by including MedISys and ProMED-mail as sources which, in combination, ranked highly during the Barboza study in numerous evaluation characteristics.

The goal of the KIWI-Zoonotic automated SMA is to reduce the number of IIPs being rated manually by users. With this in mind, it is more important for the automated SMA to maintain a low false negative rate rather than a low false positive rate. False positives are simply rated "Not Relevant" by the user community and do not become EWSs, while false negatives require more resources to locate and manually enter into the system. Since positive predictive value can be calculated as $1 - (\text{false positive rate})$ and negative predictive value can be calculated as $1 - (\text{false negative rate})$, we can alternatively say that it is more important for the automated SMA to maintain a high negative predictive value rather than a high positive predictive value.

KIWI-Zoonotic's automated SMA performed highly in specificity and negative predictive value, which is of value for our purposes. The automated SMA performed moderately in sensitivity and positive predictive value based on the expected range of 38-72% sensitivity [6]. Further

efforts to maximize SMA sensitivity, while maintaining high specificity, will benefit the KIWI system as a whole.

Conclusion

KIWI is well poised to uniquely address the challenges proposed in literature regarding event-based surveillance in the following ways: (a) KIWI allows for unique user interfaces by discipline/collaboration, the ability to “watch” individual information pieces (IIPs) and view trends of IIPs by health condition (thus allowing users to follow events from start to finish), data volume is not limited, geographic/population coverage is high/broad, and the CNPHI platform is specifically designed for multi-jurisdictional data sharing, support, and collaboration, (b) KIWI has been integrated into an already thriving community of public health professionals who discuss, comment on, and rate AIS relevancy for the verification of early warning signals, (c) KIWI utilizes a variety of sources including numerous existing event-based systems, and (d) the goal of KIWI is to support CNPHI’s existing activities in public health surveillance and response.

The automated sense-making algorithm for KIWI’s Zoonotic program is useful for the detection of IIPs related to zoonotic and emerging diseases, and it seems to perform well in maintaining a low rate of false negatives. Further evaluation would be useful in validating this over a longer duration.

The purpose of the KIWI technology is to provide situational awareness and early warning signal detection in support of surveillance activities. Resulting signals have the potential to influence public health action and complement traditional surveillance methods by providing timely information.

Limitations

The purpose of this paper was to introduce the KIWI technology, evaluate it briefly based on the KIWI-Zoonotic Pilot, and show how KIWI is uniquely designed within the context of national surveillance and collaboration. Event-based surveillance evaluation was not applied to its full capacity as such an evaluation is beyond the scope of this paper.

The KIWI technology is limited by its dependence on online sources with available RSS feeds. The timeliness of early warning signal development is limited by the timeliness of manual detection of missed signals and of user rating. There is no factor in the rating process that accounts for rater expertise per program, and average relevancy rating is currently used as a threshold for signal relevance. A limitation of this current method is that an early warning signal may be rated high by an expert and low by the majority of users and not become an early warning signal. Future work should be done to identify whether an adaptive weighted approach may correct for this current limitation.

Sample size for pilot participation in KIWI-Zoonotic was small for capturing the overall readiness of public health professionals to use such a system on a regular basis and with full integration into surveillance activities, and pilot duration was not long enough to capture seasonal patterns of disease.

Acknowledgements

We would like to acknowledge that the pilot phases were enabled through the CEDZ-IIR project which was supported by funding from the Canadian Safety and Security Program (CSSP), managed through Defence Research and Development Canada (DRDC) Centre for Security Science (CSS) and hosted by the Canadian Food Inspection Agency (CFIA) providing logistics and management for the project. Authors would like to acknowledge the CNPHI team and the CEZD-IIR project team.

References

1. Hartley DM, Nelson NP, Arthur R, Barboza P, Collier N, Lightfoot N, Linge J, Goot E, Mawudeku A, Madoff L. An overview of internet biosurveillance. *Clin Microbiol Infect* [Internet]. 2013 [cited 2016 Apr];19(11):1006-13. Available from: <http://onlinelibrary.wiley.com/doi/10.1111/1469-0691.12273/epdf>
2. Dion M, AbdelMalik P, Mawudeku A. Big data and the global public health intelligence network (GPHIN). *Can Commun Dis Rep* [Internet]. 2015 [cited 2016 Apr];41(9):209. Available from: <http://search.proquest.com/docview/1713962713?pq-origsite=gscholar>
3. Keller M, Blench M, Tolentino H, Freifeld CC, Mandl KD, Mawudeku A, Eysenbach G, Brownstein JS. Use of unstructured event-based reports for global infectious disease surveillance. *Emerg Infect Dis* [Internet]. 2009 [cited 2016 Apr];15(5):689-95. Available from: https://www.researchgate.net/profile/Clark_Freifeld/publication/24377811_Use_of_Unstructured_Event-Based_Reports_for_Global_Infectious_Disease_Surveillance/links/00b7d53565f5cc7e3000000.pdf
4. Hartley DM, Nelson NP, Walters R, Arthur R, Yangarber R, Madoff L, Linge J, Mawudeku A, Collier N, Brownstein J. The landscape of international event-based biosurveillance. *Emerg Health Threats J* [Internet]. 2010 [cited 2016 Apr];3(e3). Available from: https://www.researchgate.net/profile/Noele_Nelson/publication/223989161_Landscape_of_international_event-based_surveillance/links/0fcfd508717259f6be000000.pdf
5. Lyon A, Nunn M, Grossel G, Burgman M. Comparison of Web-Based biosecurity intelligence systems: BioCaster, EpiSPIDER and HealthMap. *Transbound Emerg Dis* [Internet]. 2012 [cited 2016 Apr];59(3):223-32. Available from: https://www.researchgate.net/profile/Geoff_Grossel/publication/51900585_Comparison_of_web-based_biosecurity_intelligence_systems_BioCaster_EpiSPIDER_and_HealthMap/links/00b4952325b4100f6ce000000.pdf

6. Barboza P, Vaillant L, Mawudeku A, Nelson NP, Hartley DM, Madoff LC, Linge JP, Collier N, Brownstein JS, Yangarber R. Evaluation of epidemic intelligence systems integrated in the early alerting and reporting project for the detection of A/H5N1 influenza events. *PLoS One* [Internet]. 2013 [cited 2016 Apr];8(3):e57252. Available from: <http://journals.plos.org/plosone/article/asset?id=10.1371%2Fjournal.pone.0057252.PDF>
7. Barboza P, Vaillant L, Le Strat Y, Hartley DM, Nelson NP, Mawudeku A, Madoff LC, Linge JP, Collier N, Brownstein JS. Factors influencing performance of internet-based biosurveillance systems used in epidemic intelligence for early detection of infectious diseases outbreaks. *PLoS One* [Internet]. 2014 [cited 2016 Apr];9(3):e90536. Available from: <http://journals.plos.org/plosone/article/asset?id=10.1371%2Fjournal.pone.0090536.PDF>
8. Velasco E, Agheneza T, Denecke K, Kirchner G, Eckmanns T. Social media and Internet-Based data in global systems for public health surveillance: A systematic review. *Milbank Q* [Internet]. 2014 [cited 2016 Apr];92(1):7-33. Available from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.436.2208&rep=rep1&type=pdf>
9. German RR, Lee L, Horan J, Milstein R, Pertowski C, Waller M. Updated guidelines for evaluating public health surveillance systems. *MMWR Recomm Rep* [Internet]. 2001 [cited 2016 Apr];50(1-35). Available from: <http://www.columbia.edu/itc/hs/pubhealth/p8475/readings/cdc-updated-guidelines.pdf>
10. Corley CD, Lancaster MJ, Brigantic RT, Chung JS, Walters RA, Arthur RR, Bruckner-Lea CJ, Calapristi A, Dowling G, Hartley DM. Assessing the continuum of event-based biosurveillance through an operational lens. *Biosecur Bioterror* [Internet]. 2012 [cited 2016 Apr];10(1):131-41. Available from: <http://online.liebertpub.com/doi/pdf/10.1089/bsp.2011.0096>
11. Mukhi SN, Aramini J, Kabani A. Contributing to communicable disease intelligence management in Canada. *Can J Infect Dis Med Microbiol* [Internet]. 2007 [cited 2016 Apr]; 18(6),353-56. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2533573/>
12. Akwar H, Kloeze H, Mukhi S. Development and validation of a standardized tool for prioritization of information sources. *Online J Public Health Inform* [internet]. 2016 [cited 2016 Oct]; 8(2), 1947-2579. Available from: <http://ojphi.org/ojs/index.php/ojphi/article/view/6720>
13. World Health Organization. *International Health Regulations (2005)* [Internet]. Switzerland: World Health Organization; 2008 [cited 2016 Apr]. 74 p. Available from: <http://www.who.int/ihr/publications/9789241596664/en/>