HAL open science

PADI-web: a multilingual event-based surveillance system for monitoring animal infectious diseases

Sarah Valentin, Elena Arsevska, Sylvain Falala, Jocelyn de Goër de Hervé, Renaud Lancelot, Alizé Mercier, Julien Rabatel, Mathieu Roche

To cite this version:

Sarah Valentin, Elena Arsevska, Sylvain Falala, Jocelyn de Goër de Hervé, Renaud Lancelot, et al.. PADI-web: a multilingual event-based surveillance system for monitoring animal infectious diseases. Computers and Electronics in Agriculture, 2020, 169, pp.105163. 10.1016/j.compag.2019.105163 . hal-02503294

HAL Id: hal-02503294

https://hal.science/hal-02503294v1

Submitted on 23 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Version of Record: https://www.sciencedirect.com/science/article/pii/S0168169919310646
Manuscript_085e135fd012cf8214acfe6697256f56

Title: PADI-web: a multilingual event-based surveillance system for monitoring animal infectious diseases

Authors

Sarah Valentin[1,2,3,*], Elena Arsevska[1,3], Sylvain Falala[1], Jocelyn de Goër[4], Renaud Lancelot[1,3], Alizé Mercier[1,3], Julien Rabatel[3], Mathieu Roche[2,3]

[1] UMR ASTRE, Univ. Montpellier, CIRAD, INRA, Montpellier, France

[2] UMR TETIS, Univ. Montpellier, AgroParisTech, CIRAD, CNRS, IRSTEA, Montpellier, France

[3] CIRAD, Montpellier, France

[4] INRA, UMR EPIA, Clermont-Ferrand, France

* Corresponding author: sarah.valentin@cirad.fr

Abstract

Global animal disease outbreak detection and monitoring rely on official sources, such as intergovernmental organisations, as well as digital media and other unofficial outlets. Manually extracting relevant information from unofficial sources is time-consuming. The Platform for Automated extraction of animal Disease Information from the web (PADI-web) is an automated biosurveillance system devoted to online news source monitoring for the detection of emerging/new animal infectious diseases by the French Epidemic Intelligence System. The tool automatically collects news via customised multilingual queries, classifies them and extracts epidemiological information. We detail each step of the PADI-web pipeline, with a focus on the new user-oriented features.

Keywords

Animal Health, Epidemic Intelligence, Web Monitoring, Text Mining

© 2019 published by Elsevier. This manuscript is made available under the Elsevier user license
https://www.elsevier.com/open-access/userlicense/1.0/

1. Introduction

Epidemic intelligence aims to detect, investigate and monitor potential health threats, while relying on formal (e.g. official notifications) and informal information sources (e.g. online news) (Kaiser et al., 2006). Combining these two types of information source has proven to enhance the performance of surveillance systems by increasing the timeliness and number of detected outbreaks, while providing further epidemiological information such as awareness or preparedness (Bahk et al., 2015; Barboza et al., 2013; Dion et al., 2015). Monitoring of unofficial sources, or so-called event-based surveillance (EBS), requires the development and implementation of systems designed to retrieve and process the daily stream of unstructured data. Several EBS systems are currently monitoring unofficial sources from the web regarding a wide range of health threats (Barboza et al., 2013), but none of them specifically focuses on animal threats. Therefore we designed the Platform for Automated extraction of animal Disease Information from the web (PADI-web[1]), an automated biosurveillance system dedicated to online news source monitoring for the detection of animal health infectious events. This tool was designed to fulfil the needs of the French Epidemic Intelligence System (FEIS, or *Veille sanitaire internationale* in French), a part of the animal health epidemiological surveillance Platform (ESA Platform[2]). Created in 2013, FEIS aims to detect and monitor animal health events threatening France at the international level using both formal sources, i.e. international animal health authorities, and informal information sources such as online media. A first version of PADI-web was previously described and evaluated (Arsevska et al., 2018). Here we present the newly released version, which includes a classification module based on a machine learning approach, a translation module to handle multilingual queries and a new interface.

[1] https://padi-web.cirad.fr/en/

[2] https://www.plateforme-esa.fr/

2. Related work

International animal health authorities, such as World Organisation for Animal Health (OIE) and Food and Agriculture Organization of the United Nations (FAO), release and store official notifications in comprehensive and structured databases. These open-access databases are available on dedicated online portals, including: the World Animal Health Information Database (WAHIS) Interface[3] and the Emergency Prevention System for Priority Animal and Plant Pests and Disease (EMPRES-i)[4]. While these organizations focus on gathering and sharing validated animal and zoonotic events, biosurveillance systems monitoring unofficial sources are not domain-specific, with most of them covering human, animal and plant diseases as well as environmental threats (Table 1). EBS systems differ with regard to the level of data processing automation. ProMED is a fully human-driven process, relying on a about 30 staff members, while other systems are fully automated (MedISys, IBIS, PADI-web) or include both automatic and manual processes (HealthMap, GPHIN). Data classification is a critical stage in EBS pipelines as it reduces the amount of irrelevant data presented to users. Automation of this process involves keyword-based or machine-learning methods such as supervised classifiers. To our knowledge, little

[3] https://www.oie.int/wahis_2/public/wahid.php/Wahidhome/Home
[4] http://empres-i.fao.org/eipws3g/

65 **Table 1. Comparison of event-based biosurveillance systems according to their scope and their data processing approaches. Systems which are no longer**
66 **operational or do not include animal health threats in their scope are not included.**

System name	ProMED	GPHIN	MedISys	HealthMap	IBIS	PADI-web
Owner	International Society of Infectious Diseases (USA)	Public Health Agency (Canada)	Joint research centre (European Union)	Boston Children's Hospital (USA)	University of Melbourne (Australia)	ESA Platform (France)
Year launched	1994	1997	2004	2006	2013	2016
Access policy	Public	Restricted	Public	Public	Public	Public
Threats covered[5]	A, H, P	A, H, P, E	A, H, P, E	A, H, P, E	A, P	A
Sources[6]	O, U	O, U	O, U	O, U	U	U
No. of languages	7	9	50	7	1	6
Translation	No	A, H	No	A	/	A
Data classification[7]	H	A (keyword-based)/ H	A (supervised classifier)	A (supervised classifier)/H	Not available	A (supervised classifier)
Interface						
Available data source[8]	Full text (NL)	Full text (EN)	Headlines (NL)	Headlines (EN/NL)	Full text (NL)	Full text (EN/NL)
Information retrieval[9]	No	TB, FB	TB, FB	FB	TB, FB	TB, FB
Spatiotemporal outputs	No	No	Temporal series, map	Temporal series, map	Map	Temporal series, map
References	(Carrion and Madoff, 2017; Velasco et al., 2014)	(Blench, 2008; Keller et al., 2009; Mykhalovskiy and Weir, 2006)	(Grishman et al., 2002; Yangarber et al., 2007)	(Brownstein et al., 2008; Freifeld et al., 2008)	(Lyon et al., 2013a, 2013b)	(Arsevska et al., 2018)

[5] A: animal, H: human, P: plant, E: environment
[6] Source types, O: official sources, U: unofficial sources
[7] H: human, A: automatic
[8] Data source language, NL: native language, EN: English-translated
[9] Retrieval from database, TB: text-based (free-text queries), FB: feature-based (epidemiological attributes)

67 information is provided about the performance of whole classification processes (Torii et al., 2011).
68 Access to source data (news article content in the native language at the source, or English-
69 translated) and the way information is retrieved from databases also differ between the different
70 tools. Most EBS systems have online portals, displaying either the whole translated content of the
71 news articles (GPHIN), or the translated headlines with a link to the original content (MedISys and
72 HealthMap). Except for HealthMap, users can retrieve news articles either via free-text queries in the
73 title or article news content (text-based retrieval) or by using epidemiological attributes such as
74 disease names or symptoms (features) previously extracted by the EBS system (feature-based
75 retrieval). ProMED alerts are disseminated by email. ProMED (Madoff, 2004), MedISys (Mantero et
76 al., 2011), HealthMap (Freifeld et al., 2008) and IBIS (Lyon et al., 2013a) are open access, while GPHIN
77 (Blench, 2008) have restricted access. PADI-web relies on a fully automated pipeline focused on
78 animal infectious diseases, capable of handling multilingual news and allowing users to perform
79 personalized queries in its database. We describe the pipeline in detail in the following sections.

80 **3. PADI-web pipeline**

81 The PADI-web pipeline involves four steps ranging from online news collection to the extraction of
82 epidemiological features.

83 **3.1. Data collection**

84 PADI-web collects news articles from Google News through customised really simple syndication
85 (RSS) feeds daily. Each RSS feed consists of an association of terms (disease names, clinical signs or
86 hosts) extracted by an integrated approach combining text mining and domain experts (Arsevska et
87 al., 2016). PADI-web currently uses two types of RSS feed. Disease-based RSS feeds include disease
88 terms (e.g. *avian flu OR avian influenza OR bird flu)* and target seven animal diseases (Table 2). These
89 diseases were selected with regard to their risk of introduction in French areas and to their
90 detrimental impacts on animal health and production chains. Moreover, to evaluate PADI-web in

91 different contexts, we targeted diseases with different epidemiological statuses (e.g. zoonotic: avian
92 influenza, endemic: foot-and-mouth disease in North Africa, emerging: African swine fever in
93 Europe), different ways of transmission (vector-borne: bluetongue, direct contact: classical swine
94 fever). Another driver for selecting diseases was their notifiable statuses (non-notifiable:
95 Schmallenberg virus[10], notifiable and enzootic in some areas, thus reported every six months: foot-
96 and-mouth disease in several African countries). Symptom-based RSS feeds consist of combinations
97 of clinical signs and hosts (e.g. *abortions AND cows*) and do not include any disease term. Those
98 feeds enable detection of diseases that are not explicitly monitored or not confirmed at the time of
99 article release, as well as unknown hazards (Arsevska et al., 2016).

100 In addition to English RSS feeds, we implemented non-English feeds to increase the local media
101 coverage. The languages were selected to target risk areas regarding specific diseases (e.g. we
102 integrated RSS feeds in Arabic for monitoring foot-and-mouth disease news in endemic countries).

103 **3.2. Data processing**

104 To avoid duplicates, PADI-web checks if retrieved articles already exist in the database based on the
105 URL. For each news article that is not a duplicate, the corresponding webpage is visited to fetch its
106 content. The title and text are cleaned to remove irrelevant elements (pictures, ads, hyperlinks, etc.),
107 using the *BeautifulSoup* python library. The *langdetect* python library is used to detect the language
108 of the article. Then all news articles that are not in English are translated using the Translator API of
109 the Microsoft Azure system. Since February 2019, the implementation of multilingual RSS feeds led
110 to a 131% increase in relevant news articles for ASF (207 in English, 272 translated), 47% for avian
111 influenza (212 in English, 99 translated) and 67% for foot-and-mouth disease (104 in English, 174
112 translated).

113 **3.3. Data classification**

[10] Although Schmallenberg virus is no longer circulating in Europe, continuous monitoring is still done.

The data classification step aims to select relevant news, which is subsequently processed by the information extraction module. By relevant news, we mean a news report that is related to a disease event (describing a current outbreak as well as prevention and control measures, preparedness, socioeconomic impacts, etc.). The new classification module uses a supervised machine learning approach and heavily relies on the scikit-learn python library. As the pipeline is generic, users can easily create new classification tasks that are independent of each other.

3.3.1. Training and classification

The training dataset is a corpus of 600 news items randomly retrieved from the PADI-web database and manually labelled by an epidemiology expert (200 relevant news articles and 400 irrelevant news articles). Data preprocessing involves the vectorization of both textual (text and title) and categorical data (source and RSS feed), followed by a feature selection step (based on a one-way ANOVA F-test). A selection of model families is trained on the dataset (random forests with various parameters, linear support vector classification, neural networks, Gaussian-based models, K-nearest neighbours, etc.), using a 5-fold cross-validation scheme. The model obtaining the highest mean accuracy score along the folds is trained on the whole dataset and is subsequently used to classify each new retrieved article. Currently Random Forest is the best classifier built with this methodology (composed of 50 trees with a maximum depth of 12), which obtains an average accuracy score of 97.2% through the 5-fold cross-validation process. We evaluated the classifier performance in a new set of 100 articles (external validation) that was preannotated by two epidemiologists (gold standard). The classifier obtained a 92% accuracy score.

3.3.2. Training dataset enhancement

The PADI-web interface allows users to label the relevance of the retrieved news articles manually. If different from the classifier label, the user label prevails over the classifier label. Each manually labelled article is added to the initial learning dataset. Therefore, each new Training Step relies on

138 the initial dataset enriched with the user contribution. It allows a quick increase in the learning
139 dataset size and adaptation of the classifier if users detect significant classification bias or errors. This
140 feature is of primary importance since we are dealing with textual data that are prone to rapid
141 change (e.g. onset of a new disease). From news articles retrieved with disease-based RSS feeds,
142 65.2% are classified as relevant by the classifier (Table 2). By comparison, only 11.3% of the news
143 articles retrieved with symptom-based RSS feeds are classified as relevant.

144 **Table 2. Number of retrieved news articles per RSS feed type between 20 Feb. 2016 and 20 Apr.**
145 **2019.** As the multilingual RSS feeds were recently added and cover a shorter period, they are not
146 included in this table.

Type of RSS feed	Number of RSS feeds	Number of retrieved articles	% of articles classified as relevant
Disease-based			
Avian influenza	2	5,393	88.3% (N = 4,764)
Foot-and-mouth disease	3	2,873	64% (N= 1,838)
African swine fever	4	2,324	50.7% (N = 1,178)
Classical swine fever	1	1,954	59.9% (N = 1,170)
Bluetongue	5	1,851	31.6% (N = 585)
Lumpy skin disease	1	505	37.6% (N = 189)
Schmallenberg virus	2	122	52.4% (N = 64)
Total	18	15,022	65.2% (N = 9,788)
Symptom-based	42	51,624	11.3% (N = 5,828)
Total	**60**	**66,646**	**23.4% (N = 15,616)**

147 **2.4. Information extraction**

148 The final step aims to extract epidemiological entities from the relevant news content. The
149 information extraction process relies on a combined method founded on rule-based systems and
150 data mining techniques, as detailed in Arsevska et al. (2018). Briefly, to extract names of diseases,
151 hosts and symptoms, we rely on a vocabulary which was created using text mining methods and
152 validated by domain experts (Arsevska et al., 2016). Locations are identified by matching the text
153 with location names from the GeoNames gazetteer (Ahlers, 2013) and dates with the rule-based
154 HeidelTime system (Strotgen and Gertz, 2010). The number of cases is extracted from a list of regular
155 expressions matching numbers in numerical or textual form. A confidence index is automatically

assigned to the extracted entities to reflect the probability that they correspond to the desired piece of epidemiological information.

4. User interface

4.1. Information retrieval

By default, the PADI-web home page displays the last 10 retrieved articles classified as relevant. The interface allows users to perform more in-depth searches into the database by combining basic and advanced filters. Users can filter news articles according to their metadata (e.g. publication date, name of the source) or their epidemiological content. The 'Keyword' field relies on a hierarchy of keywords and sub-keywords, that is for instance used to depict the relation between an entity name (e.g. "avian influenza") and variant names (e.g. "avian flu", "bird flu", etc.). For geographical entities, the keyword hierarchy is also used to consider granularity levels: a continent name (e.g. "Africa") is linked to the name of each of its countries (e.g. "Algeria" or "Nigeria"). Searching articles through the 'Keyword' form field is then very helpful as it allows users to enter a term and obtain results based on a search that will take its variants or more specific terms into account. PADI-web also allows users to type any word or multiword expression of his/her choice and to retrieve news by searching the expression in the title or text. This feature allows users to design highly personalised queries in combination with or independently from the epidemiological features.

4.2. Visualisation and annotation

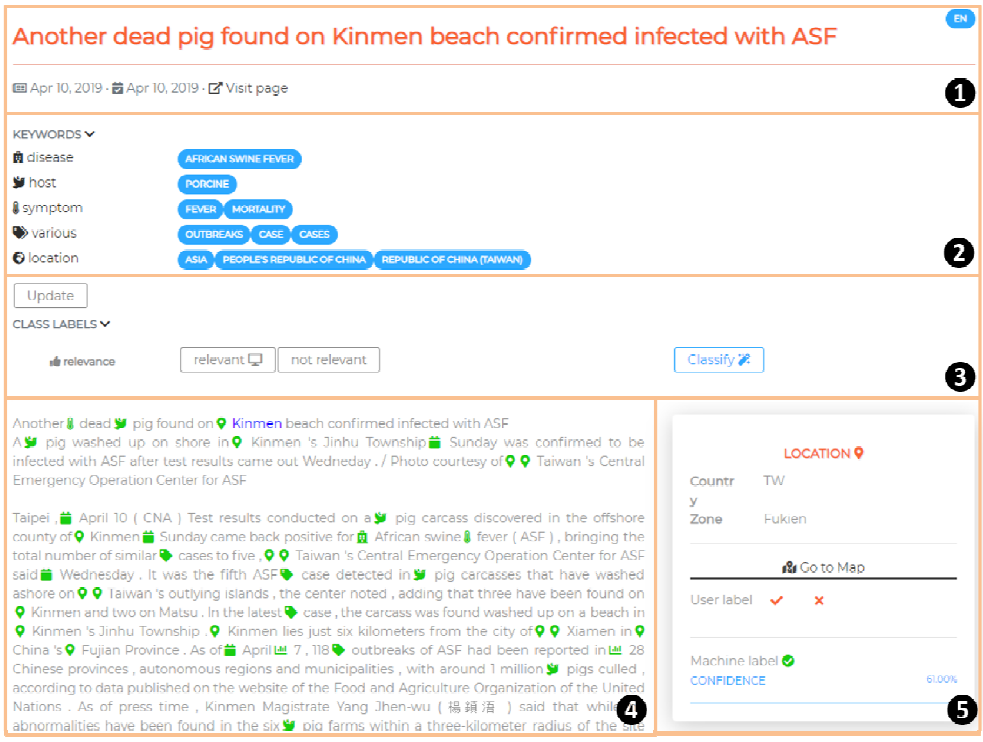
PADI-web displays the retrieved articles by visualising only their title and headlines. Translated news articles are shown in their English version, but users can also display the news in the source language. When selecting a specific news article, the PADI-web interface offers a comprehensive view combining background information about the news, i.e. title, publication date, link to the original news and source language (Figure 1 – Section 1), as well as advanced features resulting from the PADI-web process, i.e. epidemiological entities (Figure 1 – Section 2), relevance label (Figure 1 –

Section 3) and cleaned text tagged with the extracted epidemiological feature (Figure 1 – Section 4). Each type of epidemiological entity (i.e. disease, host, location, etc.) is identified with a specific icon to facilitate the reading and the detection of relevant text fragments by the user. When selecting a particular entity in the text, further details about the entity are shown, e.g. the confidence of the automatic extraction or a link to Google Map for the locations (Figure 1 – Section 5). This view also allows users to modify the automatic labels (article class and extracted entities labels).

4.3. PADI-web outputs

PADI-web offers several types of output. To frequently consult new search results outside of the application, the search results can be exported as an RSS feed. The number of news articles over time can be depicted by a histogram, using different aggregation levels (by day, month, or year). Users can export a structured dataset containing the extracted epidemiological entities from the search results and choose between different formats (CSV, json and xls).

Figure 1. View of a news report processed by PADI-web displaying: (1) basic information about the news (title, publication date, link to the original content), (2) list of epidemiological entities, (3) class label predicted by the classifier, (4) cleaned text tagged with the epidemiological entities and (5) details on the selected location 'Kinmen'.



Conclusion

In this paper, we describe PADI-web, a biosurveillance system devoted entirely to animal health surveillance. Since its first implementation, PADI-web has been substantially extended with new features to address the needs of epidemic intelligence experts in their daily monitoring activities. By providing an open-access interface, we believe that our tool can benefit a broader range of users from the animal health domain. The classification module markedly reduces the number of irrelevant news articles retrieved, thus saving considerable manual filtering time. Moreover, the module allows users to create new labels and learning datasets, thus making the classification tasks highly flexible. Multilingual queries and a translation module have been integrated to increase the media coverage. We expect that encompassing a broader range of multilingual web sources will enhance the detection of early signals in non-English speaking countries (Schwind et al., 2014). We aim to quantify those enhancements with more in-depth retrospective analysis. As future improvements, we plan to provide an event-based approach for the visualisation and analysis of PADI-web extracted information. Until now, data from official surveillance systems has been analysed in a well-established procedure, separately from unofficial data. Thus, to facilitate the comparison of official and unofficial signals, a visualization tool (EpidNews) was developed to enable visualization of both PADI-web and EMPRES-i outputs (Goel et al., 2019), and will be integrated in the PADI-web interface.

Acknowledgements

This work was funded by the French General Directorate for Food (DGAL), the French Agricultural Research Centre for International Development (CIRAD) and the SONGES Project (FEDER and Occitanie). This work was supported by the French National Research Agency (ANR) under the Investments for the Future Program (ANR-16-CONV-0004).

References

Ahlers, D., 2013. Assessment of the Accuracy of GeoNames Gazetteer Data., in: Proceedings of the 7th Workshop on Geographic Information Retrieval. Presented at the GIR'13, ACM, New York, NY, USA, pp. 74–81.

Arsevska, E., Roche, M., Hendrikx, P., Chavernac, D., Falala, S., Lancelot, R., Dufour, B., 2016. Identification of terms for detecting early signals of emerging infectious disease outbreaks on the web. Comput. Electron. Agric. 123, 104–115. https://doi.org/10.1016/j.compag.2016.02.010

Arsevska, E., Valentin, S., Rabatel, J., de Goër de Hervé, J., Falala, S., Lancelot, R., Roche, M., 2018. Web monitoring of emerging animal infectious diseases integrated in the French Animal Health Epidemic Intelligence System. PLOS ONE 13, e0199960. https://doi.org/10.1371/journal.pone.0199960

Bahk, C.Y., Scales, D.A., Mekaru, S.R., Brownstein, J.S., Freifeld, C.C., 2015. Comparing timeliness, content, and disease severity of formal and informal source outbreak reporting. BMC Infect. Dis. 15. https://doi.org/10.1186/s12879-015-0885-0

Barboza, P., Vaillant, L., Mawudeku, A., Nelson, N.P., Hartley, D.M., Madoff, L.C., Linge, J.P., Collier, N., Brownstein, J.S., Yangarber, R., Astagneau, P., on behalf of the Early Alerting, Reporting Project of the Global Health Security Initiative, 2013. Evaluation of epidemic intelligence systems integrated in the Early Alerting and Reporting project for the detection of A/H5N1 influenza events. PLoS ONE 8, e57252. https://doi.org/10.1371/journal.pone.0057252

Blench, M., 2008. Global public health intelligence network (GPHIN), in: 8th Conference of the Association for Machine Translation in the Americas. pp. 8–12.

Brownstein, J.S., Freifeld, C.C., Reis, B.Y., Mandl, K.D., 2008. Surveillance Sans Frontieres: Internet-based emerging infectious disease intelligence and the HealthMap project. PLoS Med. 5, e151.

Carrion, M., Madoff, L.C., 2017. ProMED-mail: 22 years of digital surveillance of emerging infectious diseases. Int. Health 9, 177–183. https://doi.org/10.1093/inthealth/ihx014

Dion, M., AbdelMalik, P., Mawudeku, A., 2015. Big Data and the Global Public Health Intelligence Network (GPHIN). Can. Commun. Dis. Rep. 41, 209–214.

Freifeld, C.C., Mandl, K.D., Reis, B.Y., Brownstein, J.S., 2008. HealthMap: Global Infectious Disease Monitoring through Automated Classification and Visualization of Internet Media Reports. J. Am. Med. Inform. Assoc. 15, 150–157. https://doi.org/10.1197/jamia.M2544

Goel, R., Valentin, S., Delaforge, A., Fadloun, S., Sallaberry, A., Roche, M., Poncelet, P., 2019. EpidNews: Extracting, Exploring and Annotating News for Monitoring Animal Diseases. J. Comput. Lang. 100936. https://doi.org/10.1016/j.cola.2019.100936

Grishman, R., Huttunen, S., Yangarber, R., 2002. Information extraction for enhanced access to disease outbreak reports. J. Biomed. Inform. 35, 236–246. https://doi.org/10.1016/S1532-0464(03)00013-3

Kaiser, R., Coulombier, D., Baldari, M., Morgan, D., Paquet, C., 2006. What is epidemic intelligence, and how is it being improved in Europe? Wkly. Releases 1997–2007 11, 2892. https://doi.org/10.2807/esw.11.05.02892-en

Keller, M., Blench, M., Tolentino, H., Freifeld, C.C., Mandl, K.D., Mawudeku, A., Eysenbach, G., Brownstein, J.S., 2009. Use of Unstructured Event-Based Reports for Global Infectious Disease Surveillance. Emerg. Infect. Dis. 15, 689–695. https://doi.org/10.3201/eid1505.081114

Lyon, A., Grossel, G., Burgman, M., Nunn, M., 2013a. Using internet intelligence to manage biosecurity risks: a case study for aquatic animal health. Divers. Distrib. 19, 640–650.

Lyon, A., Mooney, A., Grossel, G., 2013b. Using AquaticHealth.net to Detect Emerging Trends in Aquatic Animal Health. Agriculture 3, 299–309. https://doi.org/10.3390/agriculture3020299

Madoff, L.C., 2004. ProMED-mail: an early warning system for emerging diseases. Clin. Infect. Dis. 39, 227–232.

Mantero, J., Belyaeva, J., Linge, J., European Commission, Joint Research Centre, Institute for the Protection and the Security of the Citizen, 2011. How to maximise event-based surveillance web-systems: the example of ECDC/JRC collaboration to improve the performance of MedISys. Publications Office, Luxembourg.

Mykhalovskiy, E., Weir, L., 2006. The Global Public Health Intelligence Network and early warning outbreak detection: a Canadian contribution to global public health. Can. J. Public Health Rev. Can. Sante Publique 97, 42–44. https://doi.org/10.17269/cjph.97.756

Schwind, J.S., Wolking, D.J., Brownstein, J.S., PREDICT Consortium1, Mazet, J.A.K., Smith, W.A., 2014. Evaluation of local media surveillance for improved disease recognition and monitoring in global hotspot regions. PLoS ONE 9, e110236. https://doi.org/10.1371/journal.pone.0110236

Velasco, E., Agheneza, T., Denecke, K., Kirchner, G., Eckmanns, T., 2014. Social media and Internet-Based data in global systems for public health surveillance: A systematic review. Milbank Q. 92, 7–33.

Yangarber, R., Best, C., Von Etter, P., Fuart, F., Horby, D., Steinberger, R., 2007. Combining information about epidemic threats from multiple sources, in: Proceedings of RANLP Workshop on Multi-Source, Multilingual Information Extraction and Summarization.