



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-02503294>

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

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

3 **Authors**

4 Sarah Valentin^{1,2,3,*}, Elena Arsevska^{1,3}, Sylvain Falala¹, Jocelyn de Goër⁴, Renaud Lancelot^{1,3}, Alizé
5 Mercier^{1,3}, Julien Rabatel³, Mathieu Roche^{2,3}

6 ¹UMR ASTRE, Univ. Montpellier, CIRAD, INRA, Montpellier, France

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

8 ³CIRAD, Montpellier, France

9 ⁴INRA, UMR EPIA, Clermont-Ferrand, France

10 * Corresponding author: sarah.valentin@cirad.fr

11

12 **Abstract**

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

22

23 **Keywords**

24 Animal Health, Epidemic Intelligence, Web Monitoring, Text Mining

25

26

27 1. Introduction

28 Epidemic intelligence aims to detect, investigate and monitor potential health threats, while relying
29 on formal (e.g. official notifications) and informal information sources (e.g. online news) (Kaiser et
30 al., 2006). Combining these two types of information source has proven to enhance the performance
31 of surveillance systems by increasing the timeliness and number of detected outbreaks, while
32 providing further epidemiological information such as awareness or preparedness (Bahk et al., 2015;
33 Barboza et al., 2013; Dion et al., 2015). Monitoring of unofficial sources, or so-called event-
34 based surveillance (EBS), requires the development and implementation of systems designed to
35 retrieve and process the daily stream of unstructured data. Several EBS systems are currently
36 monitoring unofficial sources from the web regarding a wide range of health threats (Barboza et al.,
37 2013), but none of them specifically focuses on animal threats. Therefore we designed the Platform
38 for Automated extraction of animal Disease Information from the web (PADI-web¹), an automated
39 biosurveillance system dedicated to online news source monitoring for the detection of animal
40 health infectious events. This tool was designed to fulfil the needs of the French Epidemic
41 Intelligence System (FEIS, or *Veille sanitaire internationale* in French), a part of the animal health
42 epidemiological surveillance Platform (ESA Platform²). Created in 2013, FEIS aims to detect and
43 monitor animal health events threatening France at the international level using both formal sources,
44 i.e. international animal health authorities, and informal information sources such as online media. A
45 first version of PADI-web was previously described and evaluated (Arsevaska et al., 2018). Here we
46 present the newly released version, which includes a classification module based on a machine
47 learning approach, a translation module to handle multilingual queries and a new interface.

48

49

¹ <https://padi-web.cirad.fr/en/>

² <https://www.platforme-esa.fr/>

50 **2. Related work**

51 International animal health authorities, such as World Organisation for Animal Health (OIE) and Food
52 and Agriculture Organization of the United Nations (FAO), release and store official notifications in
53 comprehensive and structured databases. These open-access databases are available on dedicated
54 online portals, including: the World Animal Health Information Database (WAHIS) Interface³ and the
55 Emergency Prevention System for Priority Animal and Plant Pests and Disease (EMPRES-i)⁴. While
56 these organizations focus on gathering and sharing validated animal and zoonotic events,
57 biosurveillance systems monitoring unofficial sources are not domain-specific, with most of them
58 covering human, animal and plant diseases as well as environmental threats (Table 1). EBS systems
59 differ with regard to the level of data processing automation. ProMED is a fully human-driven
60 process, relying on a about 30 staff members, while other systems are fully automated (MediSys,
61 IBIS, PADI-web) or include both automatic and manual processes (HealthMap, GPHIN). Data
62 classification is a critical stage in EBS pipelines as it reduces the amount of irrelevant data presented
63 to users. Automation of this process involves keyword-based or machine-learning methods such as
64 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	GPIN	MedISys	HealthMap	IBIS	PADI-web
<i>Owner</i>	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)
<i>Year launched</i>	1994	1997	2004	2006	2013	2016
<i>Access policy</i>	Public	Restricted	Public	Public	Public	Public
<i>Threats covered</i> ⁵	A, H, P	A, H, P, E	A, H, P, E	A, H, P, E	A, P	A
<i>Sources</i> ⁶	O, U	O, U	O, U	O, U	U	U
<i>No. of languages</i>	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						
<i>Available data source</i> ⁸	Full text (NL)	Full text (EN)	Headlines (NL)	Headlines (EN/NL)	Full text (NL)	Full text (EN/NL)
<i>Information retrieval</i> ⁹	No	TB, FB	TB, FB	FB	TB, FB	TB, FB
<i>Spatiotemporal outputs</i>	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)	(Arsevka 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 (Arsevka 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 virus¹⁰, 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 (Arsevaska 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.

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

120 **3.3.1. Training and classification**

121 The training dataset is a corpus of 600 news items randomly retrieved from the PADI-web database
122 and manually labelled by an epidemiology expert (200 relevant news articles and 400 irrelevant news
123 articles). Data preprocessing involves the vectorization of both textual (text and title) and categorical
124 data (source and RSS feed), followed by a feature selection step (based on a one-way ANOVA F-test).
125 A selection of model families is trained on the dataset (random forests with various parameters,
126 linear support vector classification, neural networks, Gaussian-based models, K-nearest neighbours,
127 etc.), using a 5-fold cross-validation scheme. The model obtaining the highest mean accuracy score
128 along the folds is trained on the whole dataset and is subsequently used to classify each new
129 retrieved article. Currently Random Forest is the best classifier built with this methodology
130 (composed of 50 trees with a maximum depth of 12), which obtains an average accuracy score of
131 97.2% through the 5-fold cross-validation process. We evaluated the classifier performance in a new
132 set of 100 articles (external validation) that was preannotated by two epidemiologists (gold
133 standard). The classifier obtained a 92% accuracy score.

134 **3.3.2. Training dataset enhancement**

135 The PADI-web interface allows users to label the relevance of the retrieved news articles manually. If
136 different from the classifier label, the user label prevails over the classifier label. Each manually
137 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			
<i>Avian influenza</i>	2	5,393	88.3% (N = 4,764)
<i>Foot-and-mouth disease</i>	3	2,873	64% (N= 1,838)
<i>African swine fever</i>	4	2,324	50.7% (N = 1,178)
<i>Classical swine fever</i>	1	1,954	59.9% (N = 1,170)
<i>Bluetongue</i>	5	1,851	31.6% (N = 585)
<i>Lumpy skin disease</i>	1	505	37.6% (N = 189)
<i>Schmallenberg virus</i>	2	122	52.4% (N = 64)
<i>Total</i>	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

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

158 **4. User interface**

159 **4.1. Information retrieval**

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

173 **4.2. Visualisation and annotation**

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

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

186 4.3. PADI-web outputs

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

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

196

The screenshot displays the PADI-web interface for a news report. At the top, the title "Another dead pig found on Kinmen beach confirmed infected with ASF" is shown in orange, with a language selector "EN" in the top right. Below the title, the date "Apr 10, 2019" and a "Visit page" link are visible. A circled "1" highlights this header section.

The "KEYWORDS" section lists categories like "disease", "host", "symptom", "various", and "location", each with associated tags in blue boxes: "AFRICAN SWINE FEVER", "PORCINE", "FEVER", "MORTALITY", "OUTBREAKS", "CASE", "CASES", "ASIA", "PEOPLES REPUBLIC OF CHINA", and "REPUBLIC OF CHINA (TAIWAN)". A circled "2" highlights this section.

The "CLASS LABELS" section includes a "relevance" indicator, "relevant" and "not relevant" buttons, and a "Classify" button. A circled "3" highlights this section.

The main content area shows the cleaned text with epidemiological entities highlighted in green. A circled "4" highlights this text. To the right, a "LOCATION" details panel shows "Country: TW", "Zone: Fukien", a "Go to Map" button, "User label" options, and "Machine label CONFIDENCE 61.00%". A circled "5" highlights this location details panel.

197 **Conclusion**

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

214 **Acknowledgements**

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

219 **References**

220 Ahlers, D., 2013. Assessment of the Accuracy of GeoNames Gazetteer Data., in: Proceedings of the
221 7th Workshop on Geographic Information Retrieval. Presented at the GIR'13, ACM, New York, NY,
222 USA, pp. 74–81.

223 Arsevska, E., Roche, M., Hendriks, P., Chavernac, D., Falala, S., Lancelot, R., Dufour, B., 2016.
224 Identification of terms for detecting early signals of emerging infectious disease outbreaks on the
225 web. *Comput. Electron. Agric.* 123, 104–115. <https://doi.org/10.1016/j.compag.2016.02.010>
226 Arsevska, E., Valentin, S., Rabatel, J., de Goër de Hervé, J., Falala, S., Lancelot, R., Roche, M., 2018.
227 Web monitoring of emerging animal infectious diseases integrated in the French Animal Health
228 Epidemic Intelligence System. *PLOS ONE* 13, e0199960.
229 <https://doi.org/10.1371/journal.pone.0199960>
230 Bahk, C.Y., Scales, D.A., Mekaru, S.R., Brownstein, J.S., Freifeld, C.C., 2015. Comparing timeliness,
231 content, and disease severity of formal and informal source outbreak reporting. *BMC Infect. Dis.* 15.
232 <https://doi.org/10.1186/s12879-015-0885-0>
233 Barboza, P., Vaillant, L., Mawudeku, A., Nelson, N.P., Hartley, D.M., Madoff, L.C., Linge, J.P., Collier,
234 N., Brownstein, J.S., Yangarber, R., Astagneau, P., on behalf of the Early Alerting, Reporting Project of
235 the Global Health Security Initiative, 2013. Evaluation of epidemic intelligence systems integrated in
236 the Early Alerting and Reporting project for the detection of A/H5N1 influenza events. *PLoS ONE* 8,
237 e57252. <https://doi.org/10.1371/journal.pone.0057252>
238 Blench, M., 2008. Global public health intelligence network (GPHIN), in: 8th Conference of the
239 Association for Machine Translation in the Americas. pp. 8–12.
240 Brownstein, J.S., Freifeld, C.C., Reis, B.Y., Mandl, K.D., 2008. Surveillance Sans Frontieres: Internet-
241 based emerging infectious disease intelligence and the HealthMap project. *PLoS Med.* 5, e151.
242 Carrion, M., Madoff, L.C., 2017. ProMED-mail: 22 years of digital surveillance of emerging infectious
243 diseases. *Int. Health* 9, 177–183. <https://doi.org/10.1093/inthealth/ihx014>
244 Dion, M., AbdelMalik, P., Mawudeku, A., 2015. Big Data and the Global Public Health Intelligence
245 Network (GPHIN). *Can. Commun. Dis. Rep.* 41, 209–214.
246 Freifeld, C.C., Mandl, K.D., Reis, B.Y., Brownstein, J.S., 2008. HealthMap: Global Infectious Disease
247 Monitoring through Automated Classification and Visualization of Internet Media Reports. *J. Am.*
248 *Med. Inform. Assoc.* 15, 150–157. <https://doi.org/10.1197/jamia.M2544>
249 Goel, R., Valentin, S., Delaforge, A., Fadloun, S., Sallaberry, A., Roche, M., Poncelet, P., 2019.
250 EpidNews: Extracting, Exploring and Annotating News for Monitoring Animal Diseases. *J. Comput.*
251 *Lang.* 100936. <https://doi.org/10.1016/j.cola.2019.100936>
252 Grishman, R., Huttunen, S., Yangarber, R., 2002. Information extraction for enhanced access to
253 disease outbreak reports. *J. Biomed. Inform.* 35, 236–246. [https://doi.org/10.1016/S1532-](https://doi.org/10.1016/S1532-0464(03)00013-3)
254 [0464\(03\)00013-3](https://doi.org/10.1016/S1532-0464(03)00013-3)
255 Kaiser, R., Coulombier, D., Baldari, M., Morgan, D., Paquet, C., 2006. What is epidemic intelligence,
256 and how is it being improved in Europe? *Wkly. Releases 1997–2007* 11, 2892.
257 <https://doi.org/10.2807/esw.11.05.02892-en>
258 Keller, M., Blench, M., Tolentino, H., Freifeld, C.C., Mandl, K.D., Mawudeku, A., Eysenbach, G.,
259 Brownstein, J.S., 2009. Use of Unstructured Event-Based Reports for Global Infectious Disease
260 Surveillance. *Emerg. Infect. Dis.* 15, 689–695. <https://doi.org/10.3201/eid1505.081114>
261 Lyon, A., Grosse, G., Burgman, M., Nunn, M., 2013a. Using internet intelligence to manage
262 biosecurity risks: a case study for aquatic animal health. *Divers. Distrib.* 19, 640–650.
263 Lyon, A., Mooney, A., Grosse, G., 2013b. Using AquaticHealth.net to Detect Emerging Trends in
264 Aquatic Animal Health. *Agriculture* 3, 299–309. <https://doi.org/10.3390/agriculture3020299>
265 Madoff, L.C., 2004. ProMED-mail: an early warning system for emerging diseases. *Clin. Infect. Dis.* 39,
266 227–232.
267 Mantero, J., Belyaeva, J., Linge, J., European Commission, Joint Research Centre, Institute for the
268 Protection and the Security of the Citizen, 2011. How to maximise event-based surveillance web-
269 systems: the example of ECDC/JRC collaboration to improve the performance of MedISys.
270 Publications Office, Luxembourg.

271 Mykhalovskiy, E., Weir, L., 2006. The Global Public Health Intelligence Network and early warning
272 outbreak detection: a Canadian contribution to global public health. *Can. J. Public Health Rev. Can.*
273 *Sante Publique* 97, 42–44. <https://doi.org/10.17269/cjph.97.756>
274 Schwind, J.S., Wolking, D.J., Brownstein, J.S., PREDICT Consortium¹, Mazet, J.A.K., Smith, W.A., 2014.
275 Evaluation of local media surveillance for improved disease recognition and monitoring in global
276 hotspot regions. *PLoS ONE* 9, e110236. <https://doi.org/10.1371/journal.pone.0110236>
277 Velasco, E., Agheneza, T., Denecke, K., Kirchner, G., Eckmanns, T., 2014. Social media and Internet-
278 Based data in global systems for public health surveillance: A systematic review. *Milbank Q.* 92, 7–33.
279 Yangarber, R., Best, C., Von Etter, P., Fuart, F., Horby, D., Steinberger, R., 2007. Combining
280 information about epidemic threats from multiple sources, in: *Proceedings of RANLP Workshop on*
281 *Multi-Source, Multilingual Information Extraction and Summarization.*
282