ABSTRACTS

4th International Conference on Animal Health Surveillance

3 - 5 MAY 2022 · COPENHAGEN · DENMARK

www.icahs4.org

INDEX

Keynote abstracts ………………………………………………………………….. p. 2 - 7

Oral presentations ………………………………………………………………….. p. 8 - 178

Poster presentations ……………………………………………………………….. p. 179 - 278

[O2] DISSEMINATION OF INFORMATION IN EVENT-BASED SURVEILLANCE

Sarah Valentin[1,2,3], Bahdja Boudoua[4], Claire Hautefeuille[2], Renaud Lancelot[2], Mathieu Roche[1], Elena Arsevska[2]

[1]Cirad, Umr Tetis, F-34398 Montpellier, France
[2]Cirad, Umr Astre, F-34398 Montpellier, France
[3]Département de Biologie, Université de Sherbrooke, Sherbrooke, Québec, Canada
[4]INRAE, Umr Tetis, F-34398 Montpellier, France

Summary
The main role of event-based surveillance (EBS) systems is to detect signals of disease outbreaks by mining the health information published in a broad range of online formal and informal sources. In this work, we describe for the first time the flow of outbreak information between sources used by two EBS systems. We applied network analysis to highlight the value of sources in the dissemination of information on African swine fever outbreaks in 2018 and 2019. The results showed the network was highly reactive, sharing information in less than a day. Press agencies and online news sources played an essential role in detecting and sharing outbreak information from national veterinary authorities, thus bypassing the international official organisations.

Introduction
Epidemic intelligence in animal health aims to detect, investigate and monitor potential threats to animal health worldwide. While traditional indicator-based surveillance relies on structured and verified data from formal, official sources (e.g. OIE, FAO), event-based surveillance (EBS) focuses on the detection of early signals from a wide range of informal sources (e.g. social media, online news). Several EBS systems that are in use today, such as ProMED (1), HealthMap (2), and PADI-web (3), collect, process and analyse a daily stream of unstructured textual data, usually from online sources (further referred to as sources).

Several studies compared the performance of EBS systems, and their ability to detect disease outbreaks (further on referred to as events) from informal sources is recognized (4,5). However, little is known about how disease outbreak information circulates between sources before being detected by EBS systems. Knowledge on which sources communicate and disseminate outbreak-related information is crucial for understanding their role in event-based surveillance.

In this study, we assess how outbreak-related information (i.e. events) propagates between sources before being detected by an EBS system. More precisely, we aimed to answer two main questions:

- How does outbreak-related information disseminate between sources?

- What is the role of sources regarding the detection and dissemination of outbreak-related information?

We, thus, assessed the information disseminated between the sources using the network theory (6). We assessed the network based on qualitative and quantitative attributes of its components to highlight the value of sources in dissemination of outbreak-related information.

We analysed two EBS systems: HealthMap and PADI-web. HealthMap is a semi-automatic EBS system that monitors a broad range of threats, covering human, animal and plant infectious diseases as well as environmental threats. HealthMap's sources include news aggregators, social media as well as other EBS systems such as ProMED. PADI-web (Platform for Automated extraction of animal Disease Information from the web) is an automated EBS system for monitoring online news sources, developed for the needs of the French animal health epidemiological surveillance Platform (https://www.plateforme-esa.fr/). PADI-web's main source of information is the Google News aggregator (7).

Materials and Methods
To extract and analyse outbreak-related information dissemination and the roles of sources, we (i) collected the data, (ii) stored them in a database and (iii) used the latter to assess the network and the information disseminated.

Data collection

We extracted all English reports related to African swine fever (ASF) published between 1 August 2018 and 31 July 2019 from PADI-web and HealthMap databases, and containing one or several events (i.e. unverified sets of epidemiological information for an ASF case or outbreak). PADI-web relevant reports consisted of news classified as relevant. HealthMap reports included reports from different types of informal sources, e.g. online news, ProMED, Twitter, etc. We obtained 136 ASF-related reports from HealthMap and 594 ASF-related reports from PADI-web (a total of 730 reports).

Database creation

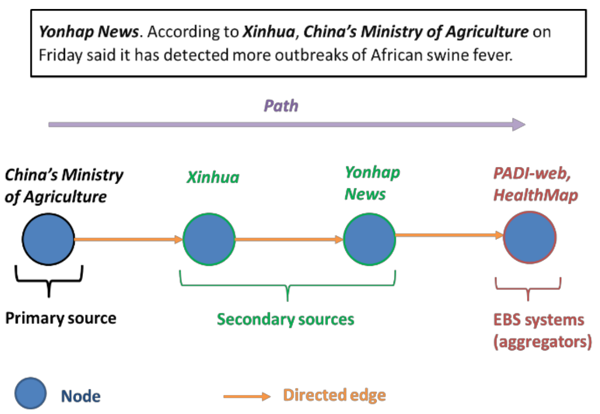
At this step, the objective was to trace back the origin of the event information. We assumed that this pathway could be deducted from the sources cited in the final reports. We distinguished the primary sources, i.e. the earliest emitter source for a given event; the intermediate sources, i.e. all sources involved in a path, except the primary source and final aggregator; and the aggregators, i.e. the EBS systems used to retrieve the final reports.
Each source was characterized by its type, i.e. international health authority (e.g. OIE), national/ local veterinary authority (e.g. Ministry of Agriculture), online news (e.g. *The New York Post*), press agency, (e.g. *Reuters*), control authority (e.g. customs), social platform (e.g. Twitter), radio/television (e.g. *Belstav.tv*), private company (e.g. food maker), animal health association (e.g. National Pig Association), and EBS systems.

Network construction and analysis

We use a graph structure to represent the network. A graph is composed of nodes and edges. Formally, a graph $G = (V; E)$ is a mathematical structure consisting of a set V of vertices (also commonly called nodes) and a set E of edges (also commonly called links) (8). The network nodes represent the sources and including the final aggregators HealthMap and PADI-web. The edges represent the dissemination of event information between two nodes (an emitter, which sends the event, and a receptor, which receives the event). We thus created an edge between an emitter S_E and receptor sources S_R if S_R cited S_E at least once. The first node of a path is called the primary node; the last node is called the final node. The primary and final nodes can be separated by intermediate nodes. The combined edges from a primary to a final node correspond to a path.

Figure 1 describes the construction of a subset of the graph based on the abovementioned example. The graph is directed, as the information is transmitted from an emitter source S_E to receptor sources S_R. A directed graph is formally defined as a graph G with each edge in E having an ordering to its vertices (i.e. such that $e_1 = (u; v)$ distinct from $e_2 = (v; u)$, for e_1; e_2 in E).
We generated two types of graphs. The event graph is specific to an event and contains the path(s) through which the event propagated. The graph shown in Figure 1 is an example of an event graph. The total graph, which is unique, consists of all paths extracted from the reports. Formally, it was generated as follows:

Figure 4. Example of network representation of information dissemination between sources. The inset contains a simplified extract from a news article containing the citation of the primary and secondary sources.



1. We initialized a graph containing all nodes (i.e. sources extracted during the database creation process), without any edges;
2. For each event P_i: for each pair of nodes (v; u) connected in the event graph of P_i, we created an edge e = (v; u) in the total graph.

Network analysis

To evaluate the network global performances regarding event dissemination, we calculated the following aggregated path and node metrics on the total graph (Table 1).

Table 2. Metrics calculated for network analysis

Metrics	Definition
Path length	Number of edges in the paths. The path length measures the number of intermediate sources between the primary source and the aggregator.
Path reactivity	Sum of the time lag of all the nodes composing the paths. The path reactivity measures the number of days between the communication by the primary source and the detection by the aggregator.
In-degree	Number of incoming edges to a node. In-degree reflects the ability of a source to collect information from different sources.
Out-degree	Number of outcoming edges from a node. Out-degree reflects the ability of a source to be cited by other sources.
Degree	Sum of in-degree and out-degree. Degree reflects the ability of a source to both collect and share information from/with different sources.

We also compared the primary and secondary nodes of each path based on their type of source. The secondary node corresponds to the node immediately following the primary node, in paths whose length is strictly higher than one edge (i.e. the first intermediate nodes).

All analyses were done using the *igraph* package available in R version 3.6 (9).

Results and discussion

HealthMap and PADI-web final reports detected 359 ASF events. The total graph contains 295 nodes and 477 edges, corresponding to 813 distinct paths. Among the paths, 47.4% (385/813) had a length of two, and 39.1% (318/813) had a length of three (Figure 36). Thus, 86.5% of the paths were transmitted from a primary node to the final node through one or two intermediate nodes. Marginally, 1.7% (14/813) of the paths had a length of one. These paths were extracted from final reports that did not cite any source. The remaining 11.8% (96/813) of the paths had a length of four or more. The network was highly reactive, with 85.7% (687/813) of the paths propagating events in less than one day.

National veterinary authorities were the primary source, representing 74% of the paths (601/813). In the remaining paths (26%, 212/813), the primary sources were online news, a local veterinary authority, a private organisation, a control authority, a press agency, a public organisation, a television channel or a radio.

Press agencies and online news were major secondary sources, representing 49% (390/799) and 32% (252/799) of the secondary sources, respectively.

International animal health organisations (FAO[1] and OIE[2]) represented only 7% (55/799) of the secondary nodes.
These results indicate that the majority of the paths of the network disseminated events originating from national authorities. The international animal health organisations (the main source of official information for international monitoring) were involved as secondary sources in fewer paths.
Importantly, this result does not reflect the intrinsic value of international animal health organisations to collect information from national authorities, as we only consider the events captured by two EBS systems. However, the results rather indicate that online news and press agencies detect information directly from primary sources, thus bypassing the international official organisations. In addition, we hypothesise that the ASF outbreaks were major events in terms of epidemiological and/or economic impact, such as the first emergence in a country or large-scale outbreak (e.g. on large pig farms, feed producers, etc.). In such cases, the event is highly shared by a broad range of sources. Online news sources, including local and regional sources, were more prone to communicate follow-up events such as an additional and local outbreak in an already affected region, thus not necessarily capturing much media interest.

Table 3. Mean (and standard deviations) of the in-degree, out-degree and degree of the sources according to their type.

	In-degree	**Out-degree**	**Degree**
Online news	1.9 (1.9)	1.5 (1.1)	3.4 (2.5)
National vet auth.	0.2 (0.5)	5.3 (9.1)	5.6 (9.5)
Radio/TV	1 (0.6)	1.2 (0.4)	2.3 (0.8)
Press agency	4.8 (7.0)	6.4 (12.1)	11.2 (18.9)
Local vet auth.	0.1 (0.2)	1.9 (2.1)	2.0 (2.1)
Private company	0.5 (0.7)	2.1 (1.9)	2.5 (2.2)
Control authority	0.2 (0.4)	2.2 (1.6)	2.4 (1.7)
Other health authority	1.0 (1.2)	2.2 (1.6)	3.2 (2.7)
Social platform	3.4 (2.3)	1.0 (0)	4.4 (2.3)
EBS system	**68.0 (80.0)**	0.3 (0.6)	**68.3 (79.6)**
International animal health organisation	6.0 (7.1)	**10.0 (12.7)**	16.0 (19.8)

EBS systems had the highest in-degree values, which was in line with the fact that HealthMap and PADI-web were the aggregators (Table 2). Excluding EBS systems, international animal health organisations had the highest degree. This indicates that, individually, international animal health organisations were connected with a large number of incoming and outcoming sources. Incoming sources were national veterinary authorities, and outcoming sources were typically online news and press agencies. Conversely, the mean degree of online news was low, indicating that their strength relies upon their number in the network. Press agencies had performances close to those obtained for international animal health organisations. Compared to online news, they had access to a broader range of information sources and, by nature, distributed their information to a vast network of sources. All the metrics had high standard deviations, thus indicating that the groups of sources were not homogeneous in terms of event dissemination, confirmed by the presence of hubs, i.e. sources with high in-degree and out-degree (6).

References

1. Carrion M et al. (2017) Int Health. 9(3):177–83.
2. Brownstein JS et al. (2008) PLoS Med. 5(7):e151.
3. Valentin S et al. (2020) Comput Electron Agric. 1;169:105163.

[1] Food and Agriculture Organisation
[2] World Organisation for Animal Health

4. Bahk CY et al. (2015) BMC Infect Dis.;15(1).
5. Barboza P et al. (2013) PLoS ONE. 5;8(3):e57252.
6. Weber MS et al. (2011) J Commun.;61(6):1062–81.
7. Arsevska E et al. (2018) PLoS ONE. 13(8):e0199960.
8. Kolaczyk ED. (2009) Statistical Analysis of Network Data: Methods and Models. p. 1–44.
9. Csardi G et al. (2006) Complex Systems:1695.

Acknowledgement
The authors thank HealthMap and PADI-web for providing the data used in this study and the members of the French Epidemic Intelligence in Animal Health for their constructive comments. 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). It was also supported by the French National Research Agency (ANR) under the Investments for the Future Program (ANR-16-CONV-0004) and by the "Monitoring outbreak events for disease surveillance in a data science context" (MOOD) project from the European Union's Horizon 2020 research and innovation program under grant agreement No. 874850 (https://mood-h2020.eu/).