Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Syndromic surveillance using web data: a systematic review

Loukas Samaras, Elena García-Barriocanal and Miguel-Angel Sicilia

Computer Science Department, University of Alcalá, Ctra. De Barcelona km. 33.6, 28871 Alcalá de Henares, Spain

2.1 Introduction: background and scope

Smart Healthcare or *Smart Health* is a new form of healthcare. This term (Active Advice, 2017) describes the technology that contributes to better diagnostic tools and better treatment for patients, by using devices and means that improve the quality of life. Smart Healthcare technologies combine the term *Smart Technology* with health, that is smart technologies used for health purposes. Such technologies (Patient@home, 2018) can be applied to health information recording from increasingly advanced electronic devices and sensors, to storing and computing this information automatically and finally to delivering the information either for personalized advice or automated actions from the collected data. Furthermore, these new technologies concern smart home services, as well as intelligent and connected medical devices.

Under these definitions, Smart Healthcare refers to smart health, including eHealth and mHealth services. The term *mHealth* is determined by the World Health Organization (2011) as the medical and public health practice supported by mobile devices. These (electronic) devices include mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices.

Syndromic surveillance, on the other hand, refers to monitoring of epidemics. According to the Centres of Disease Control and Prevention (CDC, USA, 2004) (Henning, 2004), *Syndromic surveillance has been used for early detection of outbreaks, to follow the size, spread, and tempo of outbreaks, to monitor disease trends, and to provide reassurance that an outbreak has not occurred.* Many information systems have been created based on data from the internet to fulfill this goal; the early detection of epidemics. During the last 14 years, a new approach has begun since 2004 with the famous work of Johnson et al. (2004), when the scientists tried to take advantage of the Web search queries to reveal the spread

39

and the peak of influenza in the United States. Since then, various researches have been conducted with the help of web data (social media, search engines, etc.) to establish the fundamentals of internet surveillance systems. This novel approach deals with the implementation of information systems based on data from the web to track epidemics and to create patterns and rules for an early prediction. Across all over the world, at universities, organizations, or research centers, scientists study the potential of the web in epidemics. Except from various information systems and methods created for data analysis of the web data, new terms were introduced to describe this new approach and trend, such as *Infoveillance* or *Infodemiology* (Eysenbach, 2006; Spruit & Lytras, 2018).

In this study, we examine the publications until the year of 2018. This examination was applied on an initial collection of 337 published items, of which 225 were found to be relevant. These publications consist of many different types. Most of them (86.22%) are published articles based on researches from various organizations, universities, or research centers which study the fields of syndromic surveillance or monitor epidemics. Other publications are based on dissertations, conference announcements, reviews or systematic reviews of articles, books, and one Google patent. In Fig. 2.1, we can see these types and their percentage in relation to the sum of the published items.

As we can see in Fig. 2.1, most of the publications are published articles. Nevertheless, we allocated 19 reviews or systematic reviews.

With this work we try to give a detailed analysis of the works published by identifying the most relevant ones, concerning syndromic surveillance using the Web, presenting the biggest collection of related articles ever made on this subject. By using the methodology of systematic review, there is an in-depth description and analysis of the main areas covered by these publications: the time (year) they were published, the diseases that were examined, the web data sources used by the researchers, the countries in which these diseases were
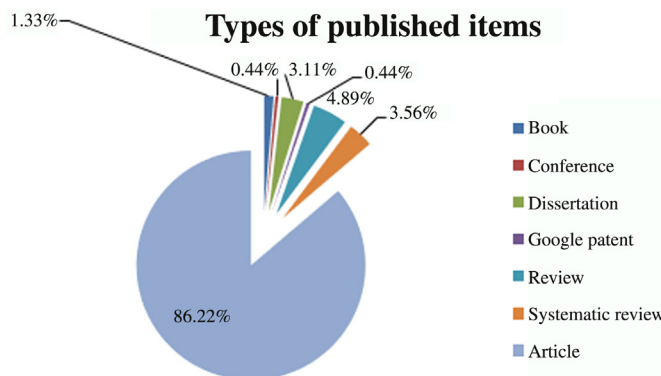

Figure 2.1
Types of all published items.

spread, the approaches of the data analysis used by the scientists, and finally, the number of scientists that have worked until now. The detailed analysis refers to 225 relevant publications that were extracted from the online database of Google Scholar (2018), which is a powerful tool for every researcher.

To conduct this review, we used a variation of the *PRISMA statements* methodology (Moher, Liberati, Tetzlaff, & Altman, 2009) and we also examine the previous reviews or systematic reviews. We identified 19 previous reviews that were conducted to analyze various articles in the previous years. Of the total reviews, most of them focus on the influenza disease or other diseases, while others examine epidemics using web data in a more general way. We consider the work of Rattanaumpawan, Boonyasiri, Vong, and Thamlikitkul (2018) as the one with the biggest collection of articles, 110 in total, without ignoring of course the good work that has been done by other scientists who conducted reviews.

Considering the above, this study is a both quantitative and qualitative analysis, and the aims of this review are:

1. To make a complete collection of articles, related to syndromic surveillance using web, available to any researcher.
2. To investigate the academic interest on this field, based on the number of published items every year. Is this growing or declining?
3. To analyze other aspects, such as the epidemic characteristics (diseases), the geographical spread of the researches and researchers, the used data, and the way the data analysis is described in these reteaches.
4. To estimate and evaluate what areas has been explored so far and what could be the possible research in the future.
5. To elaborate a novel way of systematic review on this research field that can help in deciding the importance of this research field.
6. To showcase, evaluate and align the results to the new Smart Healthcare technologies

2.2 Methodology: research protocol and stages

Our research took place from May 1 to June 30, 2018. The research model used was a variation of the PRISMA model, including a research protocol, which was created for this purpose and consists of five research stages plus a writing phase. The five stages are as follows: preparation, data retrieval, data analysis, data synthesis, and results. Below, we briefly present each stage.

2.2.1 Stage 1: Preparation, research questions, and queries

The preparation stage includes all the necessary decisions and tools to be used, such as the scope, the method designing, and the research questions. The definitions of the methods,

appropriate for the implementation of this research was critical, meaning that it should be decided at first.

Three research questions were created to examine and understand the current literature regarding syndromic surveillance using the Web. The motivation of this review is to categorize and summarize the previous work that has been done until today and find the explored and unexplored areas toward a future work. These questions are as follows:

RQ1: Is the academic interest growing or declining?
To answer this question, we must base on the current literature and examine this in a yearly basis.
RQ2: What aspects have been explored until now in the available literature?
In addition, it is useful to understand which areas are underexplored. Generally, there is a very wide field of research, since there are lots of diseases and lots of countries, which can be researched. With this question, we are seeking what has been researched and at what extend. So, the next research question should be as follows:
RQ3: What topics have been covered and which ones need further development and research in the future?
This stage was concluded with some sample data that were gathered to help in finalizing the methods and tools needed to complete this research.

The present review has gone through with the help of the search queries that were inserted in the Google Scholar search engine. The criterion for the relevance of publications is whether each one includes the following keywords in the analysis, or the variants of them: *syndromic surveillance* OR *detecting epidemics* AND *using web*.

We conducted an extensive search through Google Scholar for research papers, articles, books, dissertations, reports, and conference presentations. The main keywords (which were expected to be prevalent in the relevant papers) were *syndromic surveillance*, *predicting epidemics*, and *using web*. These keywords are generally broad terms, while some others, more specific, were used such as *Google*, *Twitter*, *Yahoo*, or *Web*. There were two final search strings (for each of the two searches made) and can be expressed in one Boolean statement as follows:

$$((A \text{ OR } B) \text{ AND } (C1 \text{ OR } C2 \text{ OR } C3 \text{ OR } C4))$$

where A, syndromic surveillance; B, predicting epidemics; C1, using Web; C2, Google; C3, Yahoo; C4, Twitter.

The search strategy contained the following design decisions:

Searched databases: Google Scholar, which contains articles from Springer Link, Science Direct, IEEE Xplore, Web of Science, etc.

Searched items: Journal articles, conference papers, workshop papers, technical reports, books reviews, and dissertations.

Search applied on: Full text, to avoid exclusion of papers that didn't include the searched keywords in abstracts or titles or used a different variant of the terms but were relevant for this review.

Language: The search was limited to papers written in English. All other languages are excluded.

Publication period: All years (2004−18).

Syndromic surveillance using web started to receive attention after 2004, as already mentioned. Therefore the chosen publication period is set to be since 2004. The outcome of the search process resulted a total of 225 relevant papers, of which 19 are relevant reviews or systematic reviews.

2.2.2 Stage 2: Data retrieval

The data were retrieved in a period of two weeks and initially included 337 published items, possible to be eligible in the review. All needed information was entered in a centralized Microsoft Excel spreadsheet for further analysis, as follows.

2.2.3 Stage 3: Data analysis: study selection and excluding criteria

The objective of the study selection was to identify papers relevant to the objectives of the review, according to the agreed scope. The search strings were set to avoid excluding relevant studies in a relatively small category or when some data were missing, for example, when there was a research using web data, but not for a specific disease. In addition, some articles were entered twice, since there were present in each one of the queries entered in the Google Scholar search engine.

2.2.4 Stage 4: Data synthesis

During this phase, extensive analysis (quantitative and qualitative) has been done regarding the information gathered from the reviewed papers and articles. This was necessary for supporting data categorization and data synthesis. Our goal was to summarize, in a quantitative way, the main areas of research related to syndromic surveillance using web data. These quantitative summaries have been included in the results section, assisted with graphs that are supplemented by the references to the included papers.

The statistical analysis has been performed in all areas concerned with the synthesis of the studies that were identified. We used multiple perspectives to indicate interesting relationships and trends within the reviewed articles, since the available literature is not limited. By doing so, we believe that a scientist can better understand the perspectives and

the trends of this type of research that has been conducted and goes into depth, finding interesting questions and views of the research field, potentially for a future work. The classification of the results includes the required characteristics to summarize, include or exclude features, to help the synthesis, and finally, the proper evaluation of all the researches included in this review.

2.2.5 Stage 5: Results analysis

This stage includes the outcomes of this research. The scope is not only to compare the previous published research to each other, but also to determine the importance of the research field of syndromic surveillance using web data.

2.2.6 Stage 6: Writing

The final stage is to write the procedures and results of this extensive research. This stage consists of all the procedures taken place during this research and we present them all in a way that helps in understanding each stage.

The entire research model can be briefly described in Fig. 2.2.

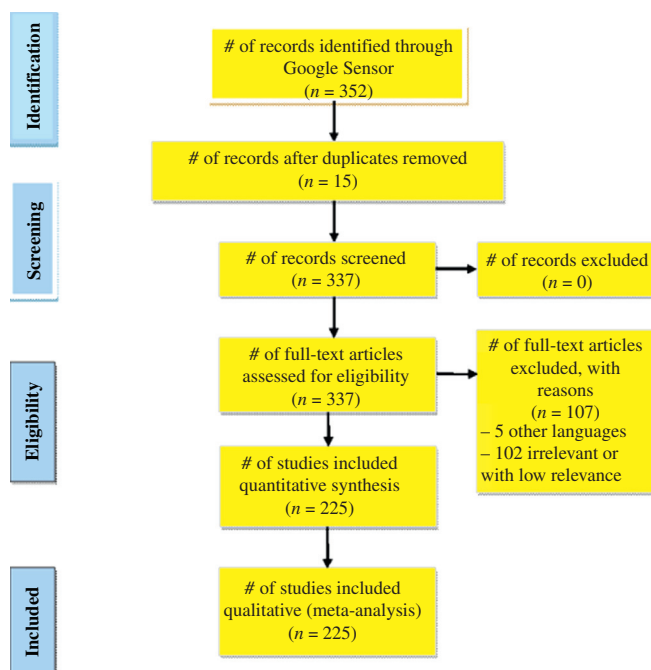**Figure 2.2**
The PRISMA research model.

2.3 Results and analysis

In this section, the results of this systematic review are presented based on the 225 papers that are relevant to the examined subject and finally selected.

2.3.1 RQ1: Is the academic interest growing or declining?

To be able to answer this question, we must see the development of the articles and other items that were published each year from 2004. That means that we must find when these publications were made and determine the trend.

The first (relevant) one is published in 2004 and concerns the study of Johnson et al. (2004), while the last one was published in 2018. Many studies occurred after 2009 as a result of the 2009 influenza pandemic. A lot of articles refer to influenza and particularly to the H1N1 virus that was responsible for many cases with severe symptoms during that year. This virus type has an unusual characteristic: it does not affect people over 60 years old but many young people, adults, and children. This virus type can lead to pneumonia if not early diagnosed and treated. The term *swine flu* was used by the news media in America and a lot of money were directed to the treatment of this disease. For instance, in Russia, the government has allocated 4 billion rubles (US$140 million) to buy the initial 43 million doses of vaccines, to perform mass swine flu vaccinations during the year 2009. Russia planned to have 35.5 million doses before the end of that year (Russian News & Information Agency, 2018). This virus had implications in many people across the world in many countries. Since 2009, a lot of relevant articles have been published, as shown in Fig. 2.3.
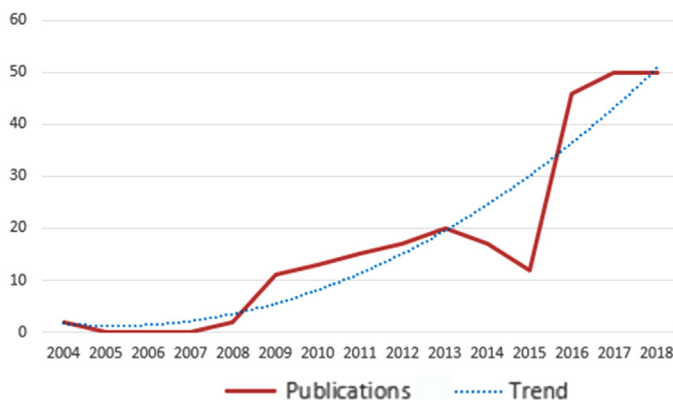

Figure 2.3.
Relevant publications by year.

In Fig. 2.3, the thicker line represents the number of publications per year, while the dotted thinner line represents the trend, as it was calculated from the following polynomial formula:

$$y = 0.3042x^2 - 1.3639x + 2.7626$$

where *y* is the expected number of publications and *x* is the current year—2003

Most articles per year are observed in 2017 (50 articles), but from 2013 until 2017, each year over 20 articles have been published regarding syndromic surveillance using the Web, except from the year 2015, in which we observe 12 publications. It seems that many scientists, research centers, or organizations appreciate the usability of the Web to provide data and tools for monitoring epidemics and outbreaks when they are effectively used along with the science of statistics. During the year 2018, we have allocated only 20 articles, but this is normal since this systematic review is conducted in June 15, 2018. As a result, the number of publications for the year 2018 is calculated as 50, using the abovementioned formula.

The results of this analysis answers to the first research question that the academic interest grows over time.

2.3.2 RQ2: Regarding syndromic surveillance using web data, what aspects have been explored until today in the available literature?

To answer this question, an analysis based on five perspectives of the current literature about the field of syndromic surveillance is conducted. The following categories of information represent the five perspectives to answer which disease, where, from what data source, with what techniques these researches have been conducted and the total number of authors of these publications, as follows in the summary:

- Disease analyzed, and health subjects used for the investigation (which disease)
- Geographical location of the studies (where)
- Web data source included in the studies (from what)
- With what statistical tools the data were handled (with what)
- How many scientists have worked to produce the published studies?

2.3.2.1 Which diseases have been explored?

Health issues are widely discussed in this research field as it is critical for modeling and predicting purposes concerning public health. Of the 225 publications, 161 (71.55%) refer explicitly to one or more health subject or disease.

Some researchers conducted researches for more than one health subject (6.22%), while others focused on one disease (65.33%). There is a relatively large percentage of researches

that did not examine specific diseases, but the syndromic surveillance generally (28.45%). Table 2.1 includes all cases.

The relative % column indicates the percentage of the publications containing at least one health subject/disease. As we can see in this table, 103 published articles are related

Table 2.1: Health subjects.

n	Diseases	Publications	%	Relative %
1	AIDS, hepatitis B, tuberculosis, influenza, and foot-and-mouth disease	1	0.44	0.62
2	Influenza and pertussis	1	0.44	0.62
3	Ebola, cholera, respiratory epidemics, syndrome coronavirus outbreak, meningitis, influenza A H7N913, chikungunya, dengue fever, mumps	1	0.44	0.62
4	African swine fever	1	0.44	0.62
5	Breast cancer	1	0.44	0.62
6	Cancer	2	0.88	1.34
7	Cholera	1	0.44	0.62
8	Cholera, dengue	1	0.44	0.62
9	Communicable diseases	1	0.44	0.62
10	Dementia	1	0.44	0.62
11	Dengue	10	4.44	6.21
12	Depression	1	0.44	0.62
13	Diseases with diverse modes of transmission (e.g., airborne droplet, vector, sexual, and fecal-oral), biology (virus, bacteria, protozoa), dengue, tuberculosis	1	0.44	0.62
14	Ebola	3	1.33	1.86
15	Ebola, Zika	1	0.44	0.62
16	Ebola, Zika, and swine flu	1	0.44	0.62
17	Foodborne illness	2	0.89	1.24
18	Hand, foot, and mouth disease epidemics	2	0.89	1.24
19	Hantavirus syndromes	1	0.44	0.62
20	HIV/AIDS	5	2.22	3.11
21	Human papillomavirus	1	0.44	0.62
22	Influenza	103	45.78	63.98
23	Influenza, cholera, rabies virus	1	0.44	0.62
24	Influenza, dengue	1	0.44	0.62
25	Influenza, tuberculosis	1	0.44	0.62
26	Listeria	1	0.44	0.62
27	Listeria, influenza	1	0.44	0.62
28	Lyme disease	1	0.44	0.62
29	Malaria	4	1.78	2.48
30	Measles	2	0.88	1.32
31	Respiratory syncytial virus	1	0.44	0.62
32	Syphilis	1	0.44	0.62
33	Systemic lupus erythematosus	1	0.44	0.62
34	Zika virus	4	1.78	2.48
35	n/a	64	28.44	

exclusively to influenza (45.78%, relative percentage 63.98%), 10 articles are about dengue, 5 articles are about HIV/AIDS, 4 articles are about malaria, 4 articles about cancer or breast cancer, and 3 articles about Ebola exclusively. There are also two references for foodborne illnesses and two for hand, foot, and mouth epidemics. Other health subjects are discussed once, such as the African swine fever, malaria, Lyme disease, listeria, foodborne illness, Ebola, depression, dementia, communicable diseases, cholera, human papillomavirus, measles, respiratory syncytial virus, syphilis, and systemic lupus erythematosus. As we see, not all the abovementioned are diseases that can be transmitted, although depression is s special case of psychological phenomenon (disorder) related to individuals. However, regarding public health, the research revealed that the Web can be a useful source of data for other nontransmitted diseases or disorders as cancer (Foroughi, Lam, Lim, Saremi, & Ahmadvand, 2016), or depression (Yang, Huang, Peng, & Tsai, 2010). There are also articles which examine more than one disease and we include them separately in the table. Such cases are influenza and listeria, influenza and pertussis, cholera and dengue, influenza, cholera and rabies virus, influenza and dengue, and influenza and tuberculosis. Three studies include more than three diseases in their area of examination.

Another aspect is the keywords used to retrieve data from the internet for those researchers who have built specific databases for this purpose. In 78 studies, the scientists used as a search term, the name of the disease, in the language of the country they examined or using the English-speaking terms. Others used the words for their symptoms or other keywords, for example, "cold" to detect influenza, based both on the weather conditions of this common disease and on social media (Shikha, Younghee, & Mihui, 2017). Weather is an important factor for the development of epidemics and we will come back to this matter, since there are some researches from scientists that are classified as relevant. We will further discuss the potential of weather forecast to help syndromic surveillance using the Web.

2.3.2.2 Where did studies take place (region, country)?

A major aspect is where the relevant published researches have been conducted, but also which regions or countries of the earth these researches concern. We may assume that there is a similarity between these two aspects, since it is usual for a study to be made in the United States, for the United States. This is only a general rule applied to most publications, but there are exceptions. For instance, a scientist in Italy conducts a research about Africa. Such a case is the study of Alicino et al. (2015), which examines the spread of Ebola virus in West Africa based on the findings of the researchers from the University of Genoa (Italy). In our case, we believe that it is more appropriate to analyze the geographical location that has been under research with the use of the internet. This is important, taking into account that the use of internet is globally available and this makes
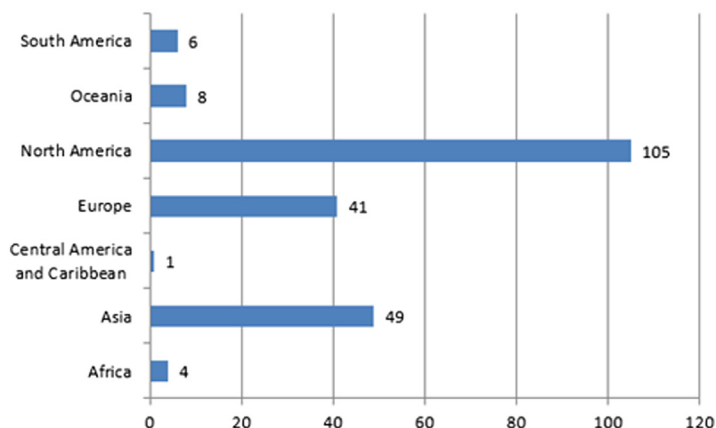
Figure 2.4
Regions examined for epidemics.

easy for data to be accessed by people in almost all countries of the world. Fig. 2.4 shows explicitly the geographic regions from which surveillance data were used.

In this figure we can see that the most explored region is North America with 105 publications. This occurs since most researchers originated from the countries of the United States and Canada, but also some other scientists have conducted researches for these countries, especially for the United States. It is important to notice that for Africa, there are only four studies conducted for the countries of West Africa (Liberia, Sierra Leone, and Guinea) for Ebola virus, since this virus is generally very common in the regions of Africa. For Europe, there are 41 publications and for Asia 49. For Asia, 17 of the articles concern China. For Oceania (Australia and New Zealand), there are eight publications, while for South America we found six publications. For Central America and Caribbean, only one publication has been made, referring to Haiti.

It is worth mentioning that eleven publications are made, examining more than one geographical region, for example, a study was made using data for Australia, Canada, Ireland, New Zealand, South Africa, the United Kingdom (England and Wales, Scotland, Northern Ireland), and the United States (Paul, Dredze, Broniatowski, & Generous, 2015). In the regional analysis, we separate the regions that were exclusively studied, while there are eleven publications concerning countries more than one. Another example is a study for Australia, Canada, and Ireland, regions quite different to each other, but they are included in one article. Other publications concern the United States and Australia or Brazil, Ukraine, Turkey, and Venezuela. We see that, in the last case, there are four countries belonging to two different geographical continents, South America and Europe. The above observations mean that the current publications concern all the continents of the world, except from Antarctica and the North Pole, since very few people live there.

Table 2.2: List of countries with publications.

n	Country	Sum	*n*	Country	Sum	*n*	Country	Sum
1	Australia	9	12	Italy	9	23	Republic of Korea	1
2	Brazil	3	13	Japan	6	24	Saudi Arabia	1
3	Canada	12	14	Korea	1	25	South Korea	5
4	China	19	15	Madagascar	1	26	Spain	4
5	Colombia	1	16	Malaysia	2	27	Sweden	3
6	Denmark	5	17	Netherlands	2	28	Taiwan	2
7	France	3	18	New Zealand	1	29	Thailand	1
8	Hong Kong	1	19	Norway	1	30	United Kingdom	18
9	India	5	20	Pakistan	1	31	United States	102
10	Indonesia	1	21	Philippines	1	32	Israel	1
11	Iran	2	22	Portugal	2			

It is also interesting to locate the countries, where the research has been conducted. All 225 researches were conducted in 32 countries. Most of the studies (102) were made in the United States. In the second place, far away, we find China with 19 publications, and the third place is taken by the United Kingdom with 18. Australia and Italy have nine publication each one. The complete list of countries is shown in Table 2.2.

From Europe, there are totally 47 publication from 9 countries on this research field. Surprisingly, Central Europe does not have too many publications, as Germany has none and France only three. In Europe, there are totally nine countries, in which researches are made: Denmark (5), France (3), Italy (9), Netherlands (2), Norway (1), Portugal (2), Spain (4), Sweden (3), and United Kingdom (18). On the contrast, it is quite interesting that countries, which are not considered as very developed or they are very small, have at least one publication, such as Madagascar, Pakistan, Philippines, and Thailand.

2.3.2.3 What is the web data source used or mentioned?

Scientists referred to or use various sources of data from the web for their experiments or/ and their analysis. We have found 169 different web sources that were used. The most popular is Twitter, as shown in Fig. 2.5.

In Fig. 2.5, we can see all the 169 web data sources. Fifty-nine publications (34.91%) have been conducted about Twitter. The next popular data source is Google Trends (30, 17.75%), which include both Google Trends and Goggle Flu Trends (GFT). Naver Trends search engine is used in one occasion. Other 18 (10.65%) researches include a combination of two or more data sources, for example, using social media and web search logs. The 15 publications (8.88%) which are mentioned as *other* include special systems developed in some countries. These systems are basically local systems used in countries to gather epidemiological data. They have been developed in specific countries and can be accessed to provide useful data and information, for example, recorded incidents by physicians,
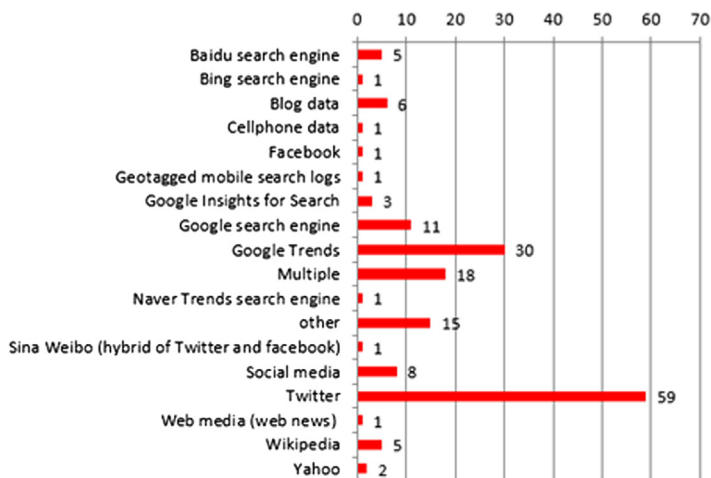
Figure 2.5
The web data sources.

cloud-based electronic health records, various other databases, or emergency medicine internet sites, such as the medical website Vårdguiden.se of Sweden (Hulth, Rydevik, & Linde, 2009).

The Google search engine follows with 11 publications (6.51%), while the Chinese Baidu search engine and Microsoft Bing were used in 6 researches (3.55%). The same number is found to blog data. Wikipedia is a web data source, used in five studies (2.96%) and data from Yahoo were used in two instances.

We can see that various web data have been used, in correlation to the syndromic surveillance data. It is important to mention that the data extraction was executed using traditional methods (use of internet sites) or by using specific API's (Application Programming Interfaces), written in programming languages, such as Python, C + +, C#, .NET, etc.

2.3.2.4 What is the method(s) used for analysis and interpretation of the data?

There are many techniques regarding the data analysis approach used by various researchers that made estimations, analysis, or predictions. We have identified 109 published items with an explicit data analysis method (48.44%). The data analysis methods or statistical approach cannot easily be categorized, as scientists use sometimes multiple methods but, as a rule, all tried to find correlations between data from the official authorities for diseases and the data obtained from the web to create prediction rules.

Twenty-four correlation techniques are used to correlate health data to data from the internet. The correlation techniques used are mentioned as Pearson correlation *R*, Spearman correlation coefficient, cross correlation function, simple correlation, or autocorrelation

function. Others use the R^2 coefficient. It must be noticed that many researchers use combinations of models, which is very common in statistics, when the probability or the validity of the results must be examined and verified in depth before they are announced. Thirty-two types of regression or autoregression models were found. These models can be summarized as follows: linear regression and the ARGO model (autoregression with Google search data) which uses publicly available online search data and not only incorporates the seasonality in influenza epidemics, but also captures changes in people's online *search behavior* over time. GARMA model is the generalized autoregression moving average model, ARIMA (Autoregressive Integrated Moving Average model) or ARMA is the Autoregressive Moving Average model, and SARIMA is the Seasonal Autoregressive Integrated Moving Average model. It must be said that in all cases, the dependent variable is the data from the official health authorities, while independent variables are the data from the sources on the internet.

Six studies have been conducted using linear, simple, and multivariate or multiple models. Seven use nonlinear models, such as logistic (logit), nonlinear regression, etc. Six researches use frequency analysis, for example, frequency of queries submitted in the internet.

It must be mentioned that there are some other models used, such as the Support Vector Machine (SVM) in two studies, and two used the Ailment Topic Aspect Model (ATAM). Of course, there are some cases that the statistical analysis requires more than one technique, for example, a research (Yanga, Santillana, & Kou, 2015) was conducted, using multiple techniques, such as:

- Auto Regression with Google search data (lag3) (ARGO),
- Root-Mean-Squared Error (RMSE),
- Mean Absolute Error (MAE), and
- Mean Absolute Percentage Error (MAPE).

Other scientists use the LASSO (Least Absolute Shrinkage and Selection Operator) models, referring to them as LASSO regression or LASSO algorithm, while others use other models, such as ANOVA, or other machine learning algorithms, or a combination of them.

Regarding an early prediction, some researchers conclude that, using web data, prediction is possible. In nine studies, early prediction time (period) varies from one day up to 6−13 weeks.

2.3.2.5 How many scientists have worked so far?

There are totally 786 authors, which contributed in the 225 publications. Some of them worked in more than one research. John S. Brownstein has the most publications, as he has participated in 17 published researches. Michael J. Paul has 10 publications as an author,

while Mark Dredze and Mauricio Santillana have nine and eight, respectively. Vasileios Lampos has seven, Elad Yom-Tov has six, and Elaine O. Nsoesie has five. Below, in Table 2.3, we see that 29 scientists have three or more publications and other 72 have two publications as authors.

Table 2.3: List of authors with more than one publication.

n	Author	Count	*n*	Author	Count	*n*	Author	Count
1	John S. Brownstein	17	36	Alina Deshpande	2	71	Loukas Samaras	2
2	Michael J. Paul	10	37	Alireza Ahmadvand	2	72	Mathieu Roche	2
3	Mark Dredze	9	38	Alok Choudhary	2	73	Megan S.C Lim	2
4	Mauricio Santillana	8	39	Andrea Freyer Dugas	2	74	Miguel-Angel Sicilia	2
5	Vasileios Lampos	7	40	Anette Hulth	2	75	Ming-Hsiang Tsou	2
6	Elad Yom-Tov	6	41	Ankit Agrawal	2	76	Mizuki Morita	2
7	Elaine O. Nsoesie	5	42	Anna C. Nagel	2	77	Naren Ramakrishnan	2
8	Ingemar J. Cox	4	43	Antonino Mazzeo	2	78	Nassim Saremi	2
9	Nicholas Generous	4	44	Avinash Gandhe	2	79	Nattiya Kanhabua	2
10	Rumi Chunara	4	45	B. Aditya Prakash	2	80	Nicola Luigi Bragazzi	2
11	Wenbiao Hu	4	46	Benjamin M. Althouse	2	81	Paolo Missier	2
12	Adam Sadilek	3	47	Benyuan Liu	2	82	Philip M. Polgreen	2
13	Aron Culotta	3	48	Brian H. Spitzberg	2	83	Qingpeng Zhang	2
14	Christina Lioma	3	49	C.R. Macintyre	2	84	Qingyu Yuan	2
15	Daniela Paolotti	3	50	Cecile Viboud	2	85	Ross Lazarus	2
16	Eiji Aramaki	3	51	Clark C. Freifeld	2	86	S.C. Kou	2
17	Eunyoung Shim	3	52	Courtney D. Corley	2	87	S.J. Yan	2
18	Gabriel J. Milinovich	3	53	Daniela Perrotta	2	88	Sachiko Maskawa	2
19	Geoffrey Colin Fairchild	3	54	Dave Osthus	2	89	Sara Romano	2
20	Gunther Eysenbach	3	55	David A. Broniatowski	2	90	Sara Y. Del Valle	2
21	Henry Kautz	3	56	Diane J. Cook	2	91	Sean D. Young	2
22	Hyekyung Woo	3	57	Diego Cedrim	2	92	Sergio Di Martino	2
23	Kåre Mølbak	3	58	Dipak K. Gupta	2	93	Seth Foldy	2
24	Lawrence C. Madoff	3	59	Elena García-Barriocanal	2	94	Shannon Rutherford	2
25	Niels Dalum Hansen	3	60	Forough Foroughi	2	95	Shihao Yang	2
26	Nigel Collier	3	61	George W. Rutherford	2	96	Simon Pollett	2
27	Reid Priedhorsky	3	62	Gustaf Rydevik	2	97	Ssu-Hsin Yu	2
28	Shilu Tong	3	63	Harshavardhan Achrekar	2	98	Sumiko R. Mekaru	2
29	Vincent Silenzio	3	64	J. Mark Gawron	2	99	Suzanne Lindsay	2
30	A.A. Chughtai	2	65	Jianfeng He	2	100	Wolfgang Nejdl	2
31	Alessandro Garcia	2	66	Jiue-An Yang	2	101	Youngtae Cho	2
32	Alessio Signorini	2	67	Josh Gray	2			
33	Alex Lamb	2	68	K. Michael Peddecord	2			
34	Alexander Romanovsky	2	69	Kumanan Wilson	2			
35	Alfred K-Y Lam	2	70	Li An	2			

2.3.3 RQ3: What topics need further development and research?

Although many aspects of the research regarding syndromic surveillance using the Web have been used and analyzed, we believe that the field of research is still very wide. Besides influenza, other diseases have been explored but not to a very wide extend.

Lately, since 2017, another aspect is under examination; the role of weather. Four studies examine the role of weather into an integrated method based on weather data to assist the epidemics prediction.

According to Van Noort, Águasa, Ballesterosb, and Gomes (2011), *a more direct approach in determining the seasonal variation of the ILI factor* p*, for example, could be the confirmation of the presence of ILI symptoms in influenza infected persons.* The above statement is important, since indicates the use of weather data to track the presence of influenza. If this is true, then future information systems could be based both on web data and weather data. Such an effort is described in the research of Shikha et al. (2017). This is a very interesting aspect, since it is not widely researched and could be researched further. A surveillance system may be created, based on weather forecasts that can be accessed through the internet and give warnings about a possible spread of a disease, for example, influenza based on weather conditions. Regarding influenza, it is considered that the spread can be extensive in areas with cold and dry weather conditions (World Health Organization, Media Center, Influenza Overview, Fact Sheet no 211, March 2003) (World Health Organization, 2003). Furthermore, it might be interesting to analyze the spread and the outbreak of diseases with discrimination between urban and rural regions, taking into account that most infectious diseases are developed in a considerable extend in crowded cases, such as big and populous areas, for example, large metropolitan cities. This could be interesting, as in rural areas, internet data may be less or more difficult to be discovered, especially in some countries with vast territories or less developed.

2.4 Discussion and conclusions

2.4.1 Results

The above results show an analysis regarding the time, the epidemics, the location, the web data, and the data analysis methods, which are presented in the examined researches and included in this review. Regarding the time, more publications are made every year based on researches from various countries from the world. This is normal, as technology and web continuously advances. The data from the web are now abundant and many scientists can take advantage of it. Many fields of science are now included in this effort to harness the web and estimate epidemics and health issues. Computer science, statistics, sociology, and medicine work together to investigate the appropriate method to exploit the web data.

Various methods are used to analyze the data in a complex framework with the ultimate goal for the social benefit. While broadband internet is nowadays widely used as the data speeds are growing with VDSL and ADSL lines (Very high speed digital subscriber line, Asymmetric digital subscriber line) by using fiber optics cables, there is still a part of the world undeveloped, such as Africa, Central America, or South America. As we saw, only four studies examine web data for Africa, one for Central America, and six for South America, as people in these countries may still have not the economy power, the expertise, or the infrastructure to widely use web services. That's why the data volume is limited in these geographical areas. On the other hand, more advanced countries seek the way of using web data for health purposes.

Regarding epidemics, influenza seems to provide plenty of data and this explains the big percentage of studies that use data for this disease. It is also a disease with many health complications and a large number of deaths annually and worldwide. Twitter is found to be used in a large scale, as it may give many and useful advantages. Through the special APIs (REST and STREAMING API), there is the capability to track messages real-time, geolocated from almost every part of the world. The messages from Twitter, called *tweets*, provide the sufficient amount of data that can be further analyzed to help in the estimation of epidemics, although this procedure requires more computer programming techniques and algorithms. Google services, on the other hand, such as the search engine or the Trends, are also extensively used and are found to be also accurate enough to track epidemics. Nevertheless, as we have conducted researches with these two popular data sources, we believe that the potential of social media and Twitter particularly is bigger, despite certain constraints and limitations.

2.4.2 Information systems and epidemics

There is a lot of discussion about the usability of internet data to track epidemics. The key focus a researcher must have under consideration is how an internet system can be managed and function as a replacement system, a supplementary system, a support system, or an extension to traditional monitoring systems. Peek, Holmes, and Sun (2014) research is about the effective management of big data, collected through both traditional and internet surveillance systems. The scientists of this research believe that data for health and biomedicine have become so big and complex that the traditional methods and tools for managing them are not efficient any more. They call this development *a big data revolution*. They examine and discover the technical aspects of big data and the infrastructure needed for the management of them, to provide useful information. They also review analytical information about health and biomedicine regarding selection of cases and control, bias and confounding in observational data and techniques for mining health dimensional data. These data come from medical records, administrative data, web search

logs, and social media. It is widely recognized that internet and electronic health data are quickly growing as technology advances. Capturing health data from various sources needs a change in the technical and management techniques regarding, such as the infrastructure, the file systems as well as the multidimensional data processing. To catch data from the internet, this can be done either with real-time techniques and nonreal-time techniques. Nevertheless, if we need early prediction of epidemics, we must support real-time data gathering and new technologies must be introduced to do so, such as Apache *Spark* and *Storm*.

Bernardo et al. (2013) examined various studies about the usability of data from the social media. Many authors claim that social media programs should primarily be used to support existing surveillance programs. The researchers of this study concluded that *the use of search queries and social media for disease surveillance are relatively recent phenomena* and there is also an evolution of the tools and the methodology to exploit them. Although these surveillance systems, based on internet data, have a support function to the traditional ones, they require a high level of familiarity about capturing social behavior through social media. The data from social media, such as Twitter, YouTube, and Facebook, are abundant, but the quality of health information among users in these media is highly variable and this may raise some concerns (Schein, Wilson, & Kealan, 2011) that social media users are exposed to unopposed viewpoints that counter core public health recommendations. We believe that this is logical but partly true as people interact with each other through social media about health issues, but it is hard to imagine that an individual could ignore the official treatment and medication or precaution measures for severe health issues just based on what people say in the social media. On the other hand, a person could be further alerted using social media.

Infoveillance (Guy, Ratzki-Leewing, Bahati, & Gwadry-Sridhar, 2011) is a term to describe the capability of real-time retrieval of internet data regarding syndromic surveillance. It is a strategy to capture real-time online data. These data can be systematically mined, aggregated, and analyzed to inform public health and policy. Social media are also considered as a real-time source of epidemic intelligence. Epidemics can be monitored through accessing the expected magnitude, peek time, and intensity, and the duration (Nsoesie, Brownstein, & Ramakrishnan, 2013). The data from the internet can inform healthcare practitioners on when to expect changes in demand for healthcare resources. A surveillance system based on real-time data could provide the means to do so. Nevertheless, in some cases, a precise estimation is not always successful. GFT is an online system that provides useful data, based on Google searches made for influenza. This system was developed by Google and during the first working period, it could also provide estimates and predictions. GFT is now no longer publishing current estimates, since it missed the emergence of the 2009 influenza pandemic and overestimated the 2012−13 influenza season epidemic (Olson, Konty, Paladini, Viboud, & Simonsen, 2013). Google searches

though were used by many scientists and researchers to track epidemics, using various statistical methods that were very successful. Generally, we must consider that a general rule cannot be found for all countries and all diseases, but each one of the latter requires deep statistical analysis; a pattern of a disease in one country may differ in another. It is possible, however, to build systems based on internet data, social media, search engines, etc., but each system requires methods, which can be similar or not. The global pattern of human behavior and globalization can help this, since internet allows people to interact to an electronic system (such as Google Search) or to each other (social media). It is true that internet-based approaches are logistically and economically appealing. However, they do not have the capacity to replace traditional surveillance systems; they should not be viewed as an alternative, but rather as an extension (Milinovich, Williams, & Hu, 2014). Of course, many people could say that the previous statement is thought to be based on today's technology. In the future, as the research continues, technology will further advance and the credibility of internet-based systems will probably be better and wide across the world. The findings of the researches until today show that the potential of creating such internet surveillance systems becomes more possible over time.

2.4.3 Impact to society, ethics, and challenges

The impact of syndromic surveillance using web data has been shown to be large, both to the academic community and to the world in general. This is revealed not only from the numerous publications made, from the large number of authors and researchers, but also from the variety of techniques and sources used. The available tools to estimate and predict epidemics or health issues are undeniably a big conquest of science by humanity. Using web data, syndromic surveillance could be accurate in a less expensive way than traditional systems. Nevertheless, traditional surveillance systems must not stop existing. Web systems could be thought as supplementary or supportive systems. Hardware and software still go on getting better. Despite the evolution of electronic systems, the role of human is still important. Moran et al. (2016) describe the role of human interaction, comparing the weather forecast models to the ones for a weather forecast and found that *epidemic forecasting is messier than weather forecasting.*

Google data and Twitter data represent almost the two-thirds (65.70%) of the total web data, used to estimate epidemics. However, we must not forget that there is significant difference between these popular tools. Using Google logs or data from other searched engines, there is the ability to track the search volume and not the search text itself, because what we really need is the aggregated number of searches, provided by these search engines. Twitter on the other side has this ability, but it first requires and it's easy enough to get the whole text which is included in a message. This means that we can read the detailed view of an individual regarding epidemics, but almost on every opinion a person

has about a topic and not just for epidemics. This is not of course a privacy violation, but it must not be used for evil purposes. Human interaction in the social media must be conceptualized as a *moment view*, which is depicted in a specific time and place through internet. This is moreover what syndromic surveillance does; an estimation of time and place.

2.4.4 Smart Healthcare innovations

Regarding modern healthcare systems, this work reveals the potential of internet syndromic surveillance systems to perform as adaptive analytic systems or as a novel research perspective in many ways, critical for Smart Healthcare: by enhancing Smart Healthcare data preprocessing and modeling and by using big data analytic techniques, model evaluation, knowledge deployment is possible through new forms of information infrastructure, as described in the work of Spruit and Lytras (2018). The key concepts applied to Smart Healthcare systems may be the following:

- How applied data science for patient-oriented healthcare can empower medical scientist and patients to more effectively and efficiently improve healthcare,
- The discipline of applied data science,
- The knowledge discovery process (KDP), and
- The meta algorithmic modeling.

Since gathering data from the internet is shown to be easy nowadays across the world, Smart Healthcare systems may drive the derived information to people in various directions: governments, scientists, medical doctors, national or international medical organizations or institutions, patients, or people who travel a lot and need direct and constant knowledge of what is happening in the world, regarding epidemics. This spread of data and information, within the framework of Smart Healthcare, could be valuable. Considering the advance of Internet of Things (IoT), the latest and future introduction of 5G networks, or smart home perspectives, internet surveillance systems could lead to a better warning and dealing with epidemics health plans and health policy decisions.

Under the above aspect, health decision policies might need a closest and greater attention to emerging technologies, such as the IoT, *cognitive computing, advanced analytics and business intelligence, 5G networks, anticipatory and context-aware computing, and advanced distributed data warehouse platforms* (Visvizi, Lytras, Damiani, & Mathkour, 2018). We believe that the key perspective would not be just the change of the technology infrastructure and social behavior, but the proper handling of health data itself for health policies.

Extending further the above into the future smart cities or villages (Lytras & Visvizi, 2018; Visvizi &Lytras, 2019) within the framework of IoT, another debate comes up regarding

data on epidemics: first, citizens' awareness of applications and solutions that are considered *smart* and secondly, their ability to use these applications and solutions. In any case, to enhance traditional surveillance systems with internet systems for epidemics, given the proper use of the health data, can lead to social-driven decisions and policies for better living conditions for humans. At the end, both syndromic surveillance using internet data and Smart Healthcare will have the impacts in society, technology, and in medical science.

2.4.5 Conclusions and outlook

The main conclusions of this review can be summarized as follows:

1. **Research**

    There is an extensive literature related to the field of syndromic surveillance using the web. This review has gathered the biggest collection of articles and publications on this field ever made, totally 225. Of course, we may assume that more publications can be found except from Google Scholar or using other search queries, but the bulk of the gathered literature in this review is quite enough to extract useful results and conclusions.

    The academic interest is still growing, especially after 2009. The years containing relevant publications extend from 2004 to the year of 2018. Since 2009, the number of publications has risen, after the 2009 pandemic of influenza. The total number of authors in publications is 786 worldwide. These two facts are important as they reveal the significance and the future perspectives of this research field.

    Many health topics have been discussed, while most of them are about influenza. It seems that for influenza a lot of data are available through internet, but the fact that totally 34 severe diseases have been included is also important, as it shows the potential of the web data to be used for syndromic surveillance.

    Most articles examine North America, the United States, Canada, Europe, and Asia, since these are mainly the origins of scientists and data are broadly available in these areas of the world. Almost all regions of the earth have been examined for epidemics, but some areas, such as Africa, South America, or Central America, are underexamined. The contribution of scientists across the world is also extensive, since researchers from 32 countries of the world have published works on this field.

2. **Internet surveillance systems**

    Twitter at first and secondly Google are the most referred and used web data sources, although other social media, search engines, health records, and web logs have been widely used. There is a rich collection of various web data and electronic systems that have been built to present an internet surveillance system.

    Various models for data analysis and techniques are used to correlate health data to data from the internet. In some occasions, scientists show that a prediction of epidemics

is possible before the official data are announced from the competent health authorities. This prediction was achieved 6−13 weeks prior to the official data. Another important aspect is that, using web data, we could estimate a disease development in shorter surveillance time intervals; while most of traditional surveillance systems monitor epidemics on weekly or monthly basis, information systems based on web data could provide daily monitoring or even for shorter periods.

3. **Smart Healthcare systems**

   Smart Healthcare innovations can use internet-based surveillance systems to assist in the awareness and direction of information to people, medical science and organizations, as well as the policy decisions. With the continuous use of internet, Global System for mobile communication networks, smart mobile devices, and health information could lead to a better understanding and informative techniques regarding epidemics. Within the concept and realization of IoT, early detection and prevention of epidemics could be the common policy for the future, rather than treatment, by reducing costs and procedures. Technology infrastructure will change to deal with the *big data revolution*, as not only humans, but also smart devices would probably be constantly connected and interact with each other.

   KDP is a constant effort to derive knowledge out of data. This can be thought as an adaptive procedure in relation to the technology progress. Health data have been always sensitive and not easy to use. The same applies for internet data, even though they are not considered as epidemical, but inspect and reveal an epidemic from other sources, for example, social media, web searches, etc. Internet, but mainly *Smart* technologies, could become irrepressible, if not treated under the previous assumption. Future research could also be dedicated on this matter. Both Internet surveillance systems and Smart Healthcare have the same destination: to improve the quality of life.

4. **Outlook**

   The potential of the research on this field remains broad and other diseases except from influenza could be further examined. In addition, the research could be driven to other aspects of syndromic surveillance, such as the weather conditions or the distinction of epidemics in between urban and rural regions with the use of data from the web. Furthermore, although the impacts of internet data for monitoring epidemics may be obvious, it could be more discussed in the future.

   Web can be a useful tool for obtaining data and make predictions for epidemics despite sometimes the restrictions of accuracy or the size of the data. Internet-based surveillance systems can be established to track epidemics and could work as alternative, supportive or as an extension to traditional systems.

   Finally, we believe that the contribution of this study is that it showcases the extended research and use of internet surveillance systems, the techniques, and the future perspectives within the concept and framework of Smart Health. Technology advance requires change of the way we think of data, information, and how to deliver them to accomplish the improvement of public health.

2.5 Teaching assignments

- Extensiveness of surveillance systems with internet data
- Data acquiring techniques and programming languages
- Social media interaction
- Early prediction and real-time monitoring of epidemics
- Smart Healthcare with the use of internet data for epidemics

Acknowledgments

Author contributions

Loukas Samaras conducted the research with the help of Dr. Miguel-Angel Sicilia and Dr.Elena García-Barriocanal. The conception and plan of the research conducted is part of Loukas Samaras' ongoing PhD work, and was done under the supervision of Miguel-Angel Sicilia and Elena García-Barriocanal.

References

Active Advice. (2017). *Smart Health − A new form of healthcare*. Available from <https://www.activeadvice.eu/news/concept-projects/what-is-smart-health-and-how-do-people-benefit/> Accessed 20.06.18.

Alicino, C., Bragazzi, N. L., Faccio, V., Amicizia, D., Panatto, D., Gasparini, R., . . . Orsi, A. (2015). Assessing Ebola-related web search behaviour: Insights and implications from an analytical study of Google Trends-based query volumes. *Infectious Diseases of Poverty*, *4*, 54. Available from https://doi.org/10.1186/s40249-015-0090-9.

Bernardo, T. M., Rajic, A., Young, I., Robiadek, K., Pham, M. T., & Funk, J. A. (2013). Scoping review on search queries and social media for disease surveillance: A chronology of innovation. *JMIR*, *15*(7). Available from https://doi.org/10.2196/jmir.2740.

Eysenbach, G. (2006). Infodemiology: Tracking flu-related searches on the web for syndromic surveillance. *AMIA Annual Symposium Proceedings Archive*, *2006*, 244−248. PMID:17238340PMCID:PMC1839505.

Foroughi, F., Lam, A. K.-Y., Lim, M. S. C., Saremi, N. A., & Ahmadvand, A. (2016). "Googling" for cancer: An infodemiological assessment of online search interests in Australia, Canada, New Zealand, the United Kingdom, and the United States. *JMIR Cancer*, *2*(1), e5. Available from https://doi.org/10.2196/cancer.5212. PMID: 28410185, PMCID: 5369660.

Google Scholar. (2018). Available from <http://scholar.google.gr/> Accessed 20.05.18.

Guy, S., Ratzki-Leewing, A., Bahati, R., & Gwadry-Sridhar, F. (2011). *Social media: A systematic review to understand the evidence and application in infodemiology*, Electronic Healthcare. eHealth 2011. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering* (Vol. 91). Berlin, Heidelberg: Springer. Available from https://doi.org/10.1007/978-3-642-29262-0_1.

Henning, K. J. (2004). Overview of syndromic surveillance. What is syndromic surveillance? *Centres of Disease Control and Prevention, Morbidity and Mortality Weekly Report (MMWR)*, *53*(Suppl), 5−11. <https://www.cdc.gov/MMWr/preview/mmwrhtml/su5301a3.htm/> Accessed 20.06.18.

Hulth, A., Rydevik, G., & Linde, A. (2009). Web queries as a source for syndromic surveillance. *PLoS ONE*, *4*(2), e4378. Available from https://doi.org/10.1371/journal.pone.0004378.

Johnson, H. A., Wagner, M. M., Hogan, W. R., Chapman, W., Olszewski, R. T., Dowling, J., & Barnas, G. (2004). Analysis of web access logs for surveillance of influenza (2004). *Medinfo*, *107*(Pt 2), 1202−1206. Available from https://doi.org/10.3233/978-1-60750-949-3-1202.

Lytras, M. D., & Visvizi, A. (2018). Who uses smart city services and what to make of it: Toward interdisciplinary smart cities research. *Sustainability*, *10*(6), 1−16. Available from https://doi.org/10.3390/su10061998.

Milinovich, G. J., Williams, G. M., & Hu, W. (2014). Internet-based surveillance systems for monitoring emerging infectious diseases. *The Lancet Infectious Diseases*, *14*(2), 160−168. Available from https://doi.org/10.1016/S1473-3099(13)70244-5.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). The PRISMA group, preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), e1000097. Available from https://doi.org/10.1371/journal.pmed.1000097.

Moran, K. R., Fairchild, G. C., Generous, N., Hickmann, K., Osthus, D., Priedhorsky, R., . . . Del Valle, S. Y. (2016). Epidemic forecasting is messier than weather forecasting: The role of human behavior and internet data streams in epidemic forecast. *The Journal of Infectious Diseases*, *214*(suppl_4), S404−S408. Available from https://doi.org/10.1093/infdis/jiw375.

Nsoesie, E. O., Brownstein, J. S., & Ramakrishnan, N. (2013). A systematic review of studies on forecasting the dynamics of Influenza outbreaks. *Influenza and Other Respiratory Viruses*, *8*(3). Available from https://doi.org/10.1111/irv.12226.

Olson, D. R., Konty, K. J., Paladini, M., Viboud, C., & Simonsen, L. (2013). Reassessing Google Flu Trends data for detection of seasonal and pandemic Influenza: A comparative epidemiological study at three geographic scales. *PLoS Computational Biology*, *9*(10), e1003256. Available from https://doi.org/10.1371/journal.pcbi.1003256.

Patient@home. (2018). Smart Helath technology: The next step in healthcare technology. Available from <https://path2025.dk/smart-health-technology/> Accessed 20.06.2018.

Paul, M. J., Dredze, M., Broniatowski, D. A., & Generous, N. (2015). Worldwide influenza surveillance through twitter. In *AAAI workshop: WWW and public health intelligence*.

Peek, N., Holmes, J. H., & Sun, J. (2014). Technical challenges for big data in biomedicine and health: Data sources, infrastructure, and analytics. *Yearbook of Medical Informatics*, *9*, 42−47. Available from https://doi.org/10.15265/IY-2014-0018.

Rattanaumpawan, P., Boonyasiri, A., Vong, S., & Thamlikitkul, V. (2018). Systematic review of electronic surveillance of infectious diseases with emphasis on antimicrobial resistance surveillance in resource-limited settings. *American Journal of Infection Control*, *46*(2), 139−146. Available from https://doi.org/10.1016/j.ajic.2017.08.006.

Russian News & Information Agency. (2018). *RIA Novosti*, Article of 20.11.2009. Available from <http://en.rian.ru/world/20091120/156921735.html/> Accessed 20.06.2018.

Schein, R., Wilson, K., & Kealan, J. (2011). *Literature review on effectiveness of the use of social media: A report for Peel Public Health*. [Region of Peel], Peel Public Health, 2011, Carleton Univercity, Ottawa, Ontario, Canada.

Shikha, V., Younghee, P., & Mihui, K. (2017). Predicting flu-rate using big data analytics based on social data and weather conditions. *Advanced Science Letters*, *23*(12), 12775−12779. Available from https://doi.org/10.1166/asl.2017.10897.

Spruit, M., & Lytras, M. D. (2018). Applied data science in patient-centric healthcare: Adaptive analytic systems for empowering physicians and patients. *Science Direct Telematics and Informatics*, *35*(4), 643−653. Available from https://doi.org/10.1016/j.tele.2018.04.002.

Van Noort, S. P., Águasa, R., Ballesterosb, S., & Gomes, M. G. (2011). The role of weather on the relation between Influenza and Influenza-like illness. *Journal of Theoretical Biology*, *298*, 131−137. Available from https://doi.org/10.1016/j.jtbi.2011.12.020.

Visvizi, A., Lytras, M. D., Damiani, E., & Mathkour, H. (2018). Policy making for smart cities: Innovation and social inclusive economic growth for sustainability. *Journal of Science and Technology Policy Management*, *9*(e:2), 126−133. Available from https://doi.org/10.1108/JSTPM-07-2018-079.

World Health Organization (2003). Media center, influenza overview, Fact Sheet no. 211 (revised March 2003). Available from <http://www.who.int/mediacentre/factsheets/2003/fs211/en/> Accessed 20.06.2018.

World Health Organization. (2011). *mHealth: New Horizons for health through mobile technologies: Based on the findings of the second global survey on eHealth (Global Observatory for eHealth Series, Volume 3). Author. ISBN 978 92 4 156425 0 (NLM classification: W 26.5).*

Yang, A. C., Huang, N. E., Peng, C.-K., & Tsai, S.-J. (2010). Do seasons have an influence on the incidence of depression? The use of an internet search engine query data as a proxy of human affect. *PLoS ONE*, *5*(10), e13728. Available from https://doi.org/10.1371/journal.pone.0013728.

Yanga, S., Santillana, M., & Kou, S. C. (2015). Accurate estimation of influenza epidemics using Google search data via ARGO. *Proceedings of the National Academy of Sciences of the United States of America*, *112* (47), 14473−14478. Available from https://doi.org/10.1073/pnas.1515373112.

Appendix: Included studies (alphabetical)

Achrekar, H., Gandhe, A., Lazarus, R., Yu, S.-H., & Liu, B. (2011). *Predicting flu trends using twitter data. International workshop on cyber-physical networking systems. Available from https://doi.org/10.1109/ INFCOMW.2011.5928903.*

Achrekar, H., Gandhe, A., Lazarus, R., Yu, S.-H., & Liu, B. (2012). *Twitter improves seasonal influenza prediction*. International conference on health informatics (HEALTHINF'12) (pp. 61−70). Vilamoura, Portugal: Nature Publishing Group.

Alessa, A., & Faezipour, M. (2018). A review of influenza detection and prediction through social networking sites. *Theoretical Biology and Medical Modelling*, *15*, 2. Available from https://doi.org/10.1186/s12976-017-0074-5.

Al-garadia, M. A., Khan, M. S., Varathan, K. D., Mujiaba, G., & Al-Kabsi, A. M. (2016). Using online social networks to track a pandemic: A systematic review. *Journal of Biomedical Informatics*, *62*, 1−11.

Alicino, C., Bragazzi, N. L., Faccio, V., Amicizia, D., Panatto, D., Gasparini, R., . . . Orsi, A. (2015). Assessing Ebola-related web search behaviour: Insights and implications from an analytical study of Google Trends-based query volumes. *Infectious Diseases of Poverty*, *4*, 54. Available from https://doi.org/10.1186/s40249-015-0090-9.

Al-Surimi, K., Khalifa, M., Bahkali, S., EL-Metwally, A., & Househ, M. (2016). The potential of social media and internet-based data in preventing and fighting infectious diseases: From internet to Twitter. In G. Rezza, & G. Ippolito (Eds.), *Emerging and re-emerging viral infections. Advances in experimental medicine and biology* (972, pp. 131−139). Cham: Springer. Available from https://doi.org/10.1007/ 5584_2016_1.

Althouse, M. B., Ng, Y. Y., & Cummings, A. T. D. (2011). Prediction of dengue incidence using search query surveillance. *PLoS Neglected Tropical Diseases*, *5*(8), e1258. Available from https://doi.org/10.1371/ journal.pntd.0001258.

Anggraeni, W., & Aristiani, L. (2016). Using Google Trend data in forecasting number of dengue fever cases with ARIMAX method case study: Surabaya, Indonesia. *Procedia Computer Science*, *124*(2017), 189−196.

Apreleva, S., & Lu, T.-C. (2017). Disease prediction system using open source data. US20170308678A1, US Application. Available from <https://patents.google.com/patent/US20170308678A1/en/> Accessed 20.06.2018.

Aramaki, E., Maskawa, S., & Morita, M. (2011). *Twitter catches the flu: Detecting influenza epidemics using Twitter. Proceedings of the conference on empirical methods in natural language processing (EMNLP'11)* (pp. 1568−1576).

Arsevska, E., Roche, M., Sylvain, F., Lancelot, R., Chavernac, D., Hendrikx, P., & Barbara, D. (2016). *Monitoring disease outbreak events on the web using text-mining approach and domain expert knowledge. International conference on language resources and evaluation 10. Portoroz, Slovénie, 23 May 2016/28 May 2016, Version publiée − Anglais.*

Aryal, A. (2016). *Developing a prototype system for syndromic surveillance and visualization using social media data* (Thesis). Rochester Institute of Technology, ProQuest Dissertations Publishing. Available from <http://scholarworks.rit.edu/cgi/viewcontent.cgi?article = 10343&context = theses/>

Aslam, A. A., Tsou, M.-H., Spitzberg, B. H., Li, A., Gawron, M., Gupta, D. K., . . . Lindsay, S. (2014). The reliability of tweets as a supplementary method of seasonal influenza surveillance. *Journal of Medical Internet Research*, *16*(11), e250. Available from https://doi.org/10.2196/jmir.3532.

Bahk, G. J., Kim, Y. S., & Park Myoung, S. (2015). Use of internet search queries to enhance surveillance of foodborne illness. *Emerging Infectious Diseases*, *21*(11), 1906−1912. Available from https://doi.org/10.3201/eid2111.141834.

Basile, L., De la Fuente, M. O., Torner, N., Martínez, A., & Jané, M. (2018). Real-time predictive seasonal influenza model in Catalonia, Spain. *PLoS ONE*, *13*(3), e0193651. Available from https://doi.org/10.1371/journal.pone.0193651.

Bernardo, T. M., Rajic, A., Young, I., Robiadek, K., Pham, M. T., & Funk, J. A. (2013). Scoping review on search queries and social media for disease surveillance: A chronology of innovation. *JMIR*, *15*(7). Available from https://doi.org/10.2196/jmir.2740.

Bhatasharrya, L., Ramashandran, A., Bhatasharrya, J., & Dogra, N. K. (2013). Google trends for formulating GIS mapping of disease outbreaks in India. *International Journal of Geoinformatics*, *9*(3), 9−19. Available from <https://www.researchgate.net/profile/Indrajit_Bhattacharya/publication/259591652_Google_Trends_for_Formulating_GIS_Mapping_of_Disease_Outbreaks_in_India/links/53d621ab0cf220632f3d86ad.pdf/>.

Bhattara, A. K. (2015). An investigation of the public health informatics research and practice in the past fifteen years from 2000 to 2014: A scoping review. In *MEDLINE, UWSpace*. Available from <http://hdl.handle.net/10012/10084/>

Bodnar, T., & Salathé, M. (2013). Validating models for disease detection using twitter. In *Companion proceedings of the 22nd international conference on world wide web (WWW'13)*, pp. 699−702. doi:10.1145/2487788.2488027.

Boit, J., & Alyami, H. (2018). Malaria surveillance system using social media. In *Association for information systems AIS electronic library (AISeL), in MWAIS 2018 proceedings*. Available from <http://aisel.aisnet.org/mwais2018/10/> Accessed 20.06.18.

Bragazzi, N. L., Barberis, I., Rosselli, R., Gianfredi, V., Nucci, D., Moretti, M., . . . Martini, M. (2017). How often people google for vaccination: Qualitative and quantitative insights from a systematic search of the web-based activities using Google Trends. In *Human vaccines & immunotherapeutics, Vol. 13, Issue 2: 10th world congress on vaccines, immunisation and immunotherapy*.

Breton, D., Bringay, S., Marques, F., Poncelet, P., & Roche, M. (2013). *Mining web data for epidemiological surveillance*, *PAKDD. Lecture Notes in Computer Science* (7769). Berlin, Heidelberg: Springer. Available from https://doi.org/10.1007/978-3-642-36778-6_2.

Briscoe, E., Appling, S., Clarkson, E., Lipskiy, N., Tyson, J., & Burkholde, J. (2017). Semantic analysis of open source data for syndromic surveillance. *Online Journal of Public Health Informatics*, *9*(1), e072. Available from https://doi.org/10.5210/ojphi.v9i1.7651.

Broniatowski, D. A., Paul, M. J., & Dredze, M. (2013). National and local influenza surveillance through Twitter: An analysis of the 2012−2013 influenza epidemic. *PLoS ONE*, *8*(12), e83672. Available from https://doi.org/10.1371/journal.pone.0083672.

Brownstein, J. S., Freifeld, C. C., & Madoff, L. C. (2009). Digital disease detection—harnessing the Web for public health surveillance. *The New England Journal of Medicine*, *360*, 2153−2157. Available from https://doi.org/10.1056/NEJMp0900702.

Byrd, K., Mansurov, A., & Baysal, O. (2016). Mining Twitter data for influenza detection and surveillance. In *Proceedings of the international workshop on software engineering in healthcare systems (SEHS'16)*, Austin, TX, pp. 43−49, May 14−22.

Dalum Hansen, N., Mølbak, K., Cox, I. J., & Lioma, C. (2017). Time-series adaptive estimation of vaccination uptake using web search queries. In *Proceedings of the 26th international conference on world wide web companion (WWW'17)*, pp. 773−774. arXiv:1702.07326

Carneiro, H. A., & Mylonakis, E. (2009). Google trends: A web-based tool for real-time surveillance of disease outbreaks. *Clinical Infectious Diseases*, *49*(10), 1557−1564. Available from https://doi.org/10.1086/630200.

Chakraborty, S., & Sub, L. (2016). Extracting signals from news streams for disease outbreak prediction. In *IEEE global conference on signal and information processing (GlobalSIP)*, Washington, DC, pp. 1300−1304. Available from https://doi.org/10.1109/GlobalSIP.2016.7906051

Chen, L., Hossain, K. S. M. T., Butler, P., Ramakrishnan, N., & Prakash, A. B. (2016). Syndromic surveillance of Flu on Twitter using weakly supervised temporal topic models. *Data Mining and Knowledge Discovery*, *30*(3), 681−710. Available from https://doi.org/10.1007/s10618-015-0434-x.

Chiu, A. P. Y., Lin, Q., & He, D. (2017). News trends and web search query of HIV/AIDS in Hong Kong. *PLoS ONE*, *12*(9), e0185004. Available from https://doi.org/10.1371/journal.pone.0185004.

Choi, J., Cho, Y., Shim, E., & Woo, H. (2016). Web-based infectious disease surveillance systems and public health perspectives: A systematic review. *BMC Public Health*, *16*(1), 1238. Available from https://doi.org/10.1186/s12889-016-3893-0.

Chunara, R., Andrews, J. R., & Brownstein, J. S. (2012). Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. *The American Journal of Tropical Medicine and Hygiene*, *86*(1), 39−45. Available from https://doi.org/10.4269/ajtmh.2012.11-0597.

Collier, N., Son, N. T., & Nguyen, N. M. (2011). OMG U got flu? Analysis of shared health messages for bio-surveillance. *Journal of Biomedical Semantics*, *2*(Suppl 5), S9. Available from https://doi.org/10.1186/2041-1480-2-S5-S9.

Corle, C. D. (2009). *Social network simulation and mining social media to advance epidemiology* (PhD thesis). University of North Texas, ProQuest Dissertations Publishing. Available from <http://digital.library.unt.edu/ark:/67531/metadc11053/m2/1/high_res_d/dissertation.pdf> Accessed 20.06.2018.

Corley, C. D., Armin, R. M., Singhy, K. P., & Cook, D. J. (2009). Monitoring influenza trends through mining social media. In *International conference on bioinformatics & computational biology*, pp. 340−346.

Corley, C. D., Cook, D. J., Mikler, A. R., & Singh, K. P. (2010). Text and structural data mining of influenza mentions in web and social media. *International Journal of Environmental Research and Public Health*, *7*(2), 596−615. Available from https://doi.org/10.3390/ijerph7020596.

Costa, F. F. (2013). Social networks, web-based tools and diseases: Implications for biomedical research. *Drug Discovery Today*, *18*(5-6), 272−281. Available from https://doi.org/10.1016/j.drudis.2012.10.006.

Coviello L., Franceschetti M., García-Herranz M. Iyad R. (2016). Predicting and containing epidemic risk using friendship networks. In Information *theory and applications wor*kshop (ITA), La Jolla, CA, pp. 1−7, Available from https://doi.org/10.1109/ITA.2016.7888201

Culotta, A. (2010a). Detecting influenza outbreaks by analyzing Twitter messages. CoRR, abs/1007.4748. Available from <http://arxiv.org/abs/1007.4748/> Accessed 20.06.18.

Culotta, A. (2010b). Towards detecting influenza outbreaks by analyzing Twitter messages. In *Proceedings of the first workshop on social media analytics (SOMA'10)*, 115−122, Washington, DC, July 25−28. ACM, New York, NY ©2010, table of contents. ISBN: 978-1-4503-0217-3.

Culotta, A. (2012). Lightweight methods to estimate Influenza rates and alcohol sales volume from Twitter messages. *Language Resources and Evaluation*, *47*(1), 217−238. Available from https://doi.org/10.1007/s10579-012-9185-0.

Curtis, S., Fair, A., Wistow, J., Val, D. V., & Oven, K. (2017). Impact of extreme weather events and climate change for health and social care systems. *Environmental Health*, *16*(Suppl 1), 128. Available from https://doi.org/10.1186/s12940-017-0324-3.

D'Avanzo, E., Pilato, G., & Lytras, M. D. (2017). Using Twitter sentiment and emotions analysis of Google Trends for decisions making. *Program*, *51*(3), 332−350. Available from https://doi.org/10.1108/PROG-02-2016-0015.

Dalton, C., Carlson, S., Hons, B., Butler, M., Cassano, D., Clarke, S., . . . Durrheim, D. (2017). Insights from flutracking: Thirteen tips to growing a web-based participatory surveillance system. *JMIR Public Health and Surveillance*, *3*(3), e48. Available from https://doi.org/10.2196/publichealth.7333.

Deiner, M. S., Fathy, C., Kim, J., Niemeyer, K., Ramirez, D., Ackley, S. F., . . . Porco, T. C. (2017). Facebook and Twitter vaccine sentiment in response to measles outbreaks. *Health Informatics Journal*. Available from https://doi.org/10.1177/1460458217740723.

Doan, S., Ohno-Machado, L., & Collier, N. (2012). Enhancing Twitter data analysis with simple semantic filtering: Example in tracking Influenza-like illnesses. In *IEEE HISB 2012 conference*, La Jolla, CA, September 27−28. USarXiv:1210.0848v1 [cs.SI].

Dredze, M., Paul, M. J., Bergsma, S., & Tran, H. (2012). Carmen: A twitter geolocation system with applications to public health. In *AAAI workshop on expanding the boundaries of health informatics using artificial intelligence (HIAI)*. Available from <https://pdfs.semanticscholar.org/9bc4/6fb12f2c7fae0e9e56e734e6efb9ca07fd98.pdf/> Accessed 20.06.18.

Dugas, A. F., Hsieh, Y.-H., Levin, S. R., Pines, J. M., Mareiniss, D. P., Mohareb, A., . . . Rothman, R. E. (2012). Google Flu Trends: Correlation with emergency department Influenza rates and crowding metrics. *Clinical Infectious Diseases*, *54*(4), 463−469. Available from https://doi.org/10.1093/cid/cir883.

Dugas, A. F., Jalalpour, M., Gel, Y., Levin, S., Torcaso, F., Igusa, T., & Rothman, R. E. (2013). Influenza forecasting with Google flu trends. *PLoS ONE*, *8*(2), e56176. Available from https://doi.org/10.1371/journal.pone.0056176.

Elkin, L. S., Topal, K., & Bebek, G. (2017). Network based model of social media big data predicts contagious disease diffusion. *Information Discovery and Delivery*, *45*(3), 110−120. Available from https://doi.org/10.1108/IDD-05-2017-0046.

Eysenbach, G. (2006). Infodemiology: Tracking flu-related searches on the web for syndromic surveillance. In *AMIA annual symposium proceedings*, pp. 244−248.

Eysenbach, G. (2009). Infodemiology and infoveillance: Framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. *Journal of Medical Internet Research*, *11*(1), e11. Available from https://doi.org/10.2196/jmir.1157.

Eysenbach, G. (2016). Impact of predicting health care utilization via web search behavior: A data-driven analysis. *Journal of Medical Internet Research*, *18*(9), e251. Available from https://doi.org/10.2196/jmir.6240.

Fairc, G. C. (2014). *Improving disease surveillance: Sentinel surveillance network design and novel uses of Wikipedia* (PhD thesis). The University of Iowa, ProQuest Dissertations Publishing. Available from <https://search.proquest.com/openview/df82c68765800af587dbf53ef6d8e8a7/1?pq-origsite = gscholar&cbl = 18750&diss = y/> Accessed 20.06.18.

Fleming, E., & Pamelá, P. (2014). *Using social media as a method for early indications & warnings of biological threats*. Capstone Project, Florida International University, Dr. Stack, April 9, 2014. Available from <http://maga.fiu.edu/academic-tracks/capstone-project/2014-capstone-working-papers/southcom-pam_eric_social-media-and-biosurveillance-capstone-april-2014-edited-by-dawndavies.pdf > .

Foldy, S., Foldy, S., Biedrzycki, P. A., Barthell, E. N., Healy-Haney, N., Baker, B. K., . . . Pemble, KimR. (2004). Syndromic surveillance using regional emergency medicine internet. *Annals of Emergency Medicine*, *44*(3), 242−246. Available from https://doi.org/10.1016/j.annemergmed.2004.01.019.

Foroughi, F., Lam, A. K.-Y., Lim, M. S. C., Saremi, N., & Ahmadvand, A. (2016). "Googling" for cancer: An infodemiological assessment of online search interests in Australia, Canada, New Zealand, the United Kingdom, and the United States. *JMIR Cancer*, *2*(1), e5. Available from https://doi.org/10.2196/cancer.5212.

Friesema, I. H. M., Koppeschaar, C. E., Donker, G. A., Dijkstra, F., Van Noort, S. P., Smallenburg, R., . . . van der Sande, M. A. B. (2009). Internet-based monitoring of Influenza-like illness in the general population: Experience of five Influenza seasons in The Netherlands. *Vaccine*, *27*(45), 6353−6357. Available from https://doi.org/10.1016/j.vaccine.2009.05.042.

Generous, N., Fairchild, G., Deshpande, A., Del Valle, S. Y., & Priedhorsky, R. (2014). Global disease monitoring and forecasting with Wikipedia. *PLoS Computational Biology*, *10*(11), e1003892. Available from https://doi.org/10.1371/journal.pcbi.1003892.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*, 1012−1014. Available from https://doi.org/10.1038/nature07634.

Girond, F., Randrianasolo, L., Randriamampionona, L., Rakotomanana, F., Randrianarivelojosia, M., Ratsitorahina, M., . . . Piola, P. (2017). Analysing trends and forecasting malaria epidemics in Madagascar

using a sentinel surveillance network: A web-based application. *Malaria Journal*, *16*. Available from https://doi.org/10.1186/s12936-017-1728-9, 72.

Guo, P., Zhang, J., Wang, L., Yang, S., Luo, G., Deng, C., . . . Zhang, Q. (2017). Monitoring seasonal influenza epidemics by using internet search data with an ensemble penalized regression model. *Scientific Reports*, *7*. Available from https://doi.org/10.1038/srep46469, Article number: 46469.

Guy, S., Ratzki-Leewing, A., Bahati, R., & Gwadry-Sridha, F. (2011). *Social media: A systematic review to understand the evidence and application in infodemiology*, Electronic Healthcare. eHealth 2011. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering* (Vol. 91). Berlin, Heidelberg: Springer. Available from https://doi.org/10.1007/978-3-642-29262-0_1.

Hansen, N. D., Lioma, C., & Mølbak, K. (2016). Ensemble learned vaccination uptake prediction using web search queries. In *Proceedings of the 25th ACM international on conference on information and knowledge management (CIKM'16)*, pp. 1953−1956.

Hansen, N. D., Mølbak, K., Cox, I. J., & Lioma, C. (2018). Predicting antimicrobial drug consumption using web search data. In *Proceedings of the 2018 international conference on digital health (DH'18)*, pp. 133−142.

Hart, M., Stetten, N., Islam, S., & Pizarro, K. (2017). Twitter and Public Health (Part 2): Qualitative analysis of how individual health professionals outside organizations use microblogging to promote and disseminate health-related information. *JMIR Public Health and Surveillance*, *3*(4), e54.

Hartley, D., Nelson, N., Walters, R., Arthur, R., Yangarber, R., Madoff, L., . . . Lightfoot, N. (2010). The landscape of international event-based biosurveillance. *Threats Journal*, *3*(1), 7096. Available from https://doi.org/10.3402/ehtj.v3i0.7096.

Hartley, D. M., Giannini, C. M., Wilson, S., Frieder, O., Margolis, P. A., Kotagal, U. R., . . . Macaluso, M. (2017). Coughing, sneezing, and aching online: Twitter and the volume of influenza-like illness in a pediatric hospital. *PLoS ONE*, *12*(7), e0182008. Available from https://doi.org/10.1371/journal.pone.0182008.

Harvey, D. (2010). Informatics research proposal predicting influenza trends from blogspsu.edu. doi: 10.1.1.577.12.

Hill, S., Mao, J., Ungar, L., Hennessy, S., Leonard, C. E., & Holmes, J. (2011). Natural supplements for H1N1 Influenza: Retrospective observational infodemiology study of information and search activity on the Internet. *Journal of Medical Internet Research*, *13*(2), e36.

Hirose, H., & Wang, L. (2012). Prediction of infectious disease spread using Twitter: A case of Influenza. *The 5th international symposium on parallel architectures*, Vol. 100. Available from https://doi.org/10.1109/PAAP.2012.23.

Ho, C. Ching, Ting, C.-Y., & Raja, D. B. (2018). Using public open data to predict dengue epidemic: Assessment of weather variability, population density, and land use as predictor variables for dengue outbreak prediction using support vector machine. *Indian Journal of Science & Technology*, *11*(4). Available from https://doi.org/10.17485/ijst/2018/v11i4/115405.

Huang, D.-C., & Wang, J.-F. (2018). Monitoring hand, foot and mouth disease by combining search engine query data and meteorological factors. *Science of the Total Environment*, *612*, 1293−1299. Available from https://doi.org/10.1016/j.scitotenv.2017.09.01.

Huang, J., Zhao, H., & Zhang J. (2013). Detecting flu transmission by social sensor in China. *In Green computing and communications (GreenCom), In 2013 IEEE and internet of things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing*, IEEE. Available from https://doi.org/10.1109/GreenCom-iThings-CPSCom.2013.216.

Hulth, A., & Rydevik, G. (2011). GET WELL: An automated surveillance system for gaining new epidemiological knowledge. *BMC Public Health*, *11*, 252. Available from https://doi.org/10.1186/1471-2458-11-252, Published: 21 April 2011.

Hulth, A., Rydevik, G., & Linde Annika. (2009). Web queries as a source for syndromic surveillance. *PLoS ONE.*, *4*(2), e4378. Available from https://doi.org/10.1371/journal.pone.0004378.

Hwang, S., Clarite, D. S., Elijorde, F. I., Gerardo, B. D., & Byun, Y. (2016). A web-based analysis for dengue tracking and prediction using artificial neural network. In *Advanced science and technology letters*. Science & Engineering Research Support Society: Sandy Bay, TAS, Australia, Vol. 122, pp. 160−164.

Iso, H., Wakamiya, S., & Aramaki, E. (2016). Forecasting word model: Twitter-based influenza surveillance and prediction. In *Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical papers*, Osaka, Japan, December 11−17, pp. 76−86.

Jadhav, A., Andrews, D., Fiksdal, A., Kumbamu, A., McCormick, J. B., Misitano, A., . . . Pathak, J. (2014). Comparative analysis of online health queries originating from personal computers and smart devices on a consumer health information portal. *Journal of Medical Internet Research*, *16*(7), e160. Available from https://doi.org/10.2196/jmir.3186.

Jena, A. B., Mandic, P. K., Weaver, L., & Seabury, S. A. (2013). Predicting new diagnoses of HIV infection using internet search engine data. *Clinical Infectious Diseases*, *56*(9), 1352−1353. Available from https://doi.org/10.1093/cid/cit022.

Ji, X., Chun, S. A., & Gelle, J. (2012). Epidemic outbreak and spread detection system based on twitter data. In *International conference on health information science*, Vol. 7231. Springer, Berlin, Heidelberg. Available from https://doi.org/10.1007/978-3-642-29361-0_19, Online ISBN 978-3-642-29361-0.

Johnson, H. A., Wagner, M. M., Hogan, W. R., Chapman, W., Olszewski, R. T., Dowling, J., & Barnas, G. (2004). Analysis of web access logs for surveillance of influenza. *Medinfo*, *107*(Pt 2), 1202−1206. Available from https://doi.org/10.3233/978-1-60750-949-3-1202.

Kandula, S., Hsu, D., & Shaman, J. (2017). Subregional nowcasts of seasonal influenza using search trends. *Journal of Medical Internet Research*, *19*(11), e370. Available from https://doi.org/10.2196/jmir.7486.

Kang, M., Zhong, H., He, J., Rutherford, S., & Yang, F. (2013). Using google trends for Influenza surveillance in South China. *PLoS ONE*, *8*(1), e55205. Available from https://doi.org/10.1371/journal.pone.0055205.

Kangbai, J. B. (2016). Social network analysis and modeling of cellphone-based syndromic surveillance data for Ebola in Sierra Leone. *Asian Pacific Journal of Tropical Medicine*, *9*(9), 851−855. Available from https://doi.org/10.1016/j.apjtm.2016.07.005.

Kazemi, D. M., Borsari, B., Levine, M. J., & Dooley, B. (2017). Systematic review of surveillance by social media platforms for illicit drug use. *Journal of Public Health*, *39*(4), 763−776. Available from https://doi.org/10.1093/pubmed/fdx020.

Khatua, A., & Khatua, A. (2016). Immediate and long-term effects of 2016 Zika outbreak: A twitter-based study. In *2016 IEEE 18th international conference on e-Health Networking, Applications and Services (Healthcom)*, Munich, pp. 1−6. Available from https://doi.org/10.1109/HealthCom.2016.7749496.

Lamb, A., Paul, M. J., & Dredze, M. (2012). Investigating Twitter as a source for studying behavioral responses to epidemics. In *AAAI fall symposium: Information retrieval and knowledge discovery in biomedical text*, Arlington, VA, pp. 81−83.

Lamb, A., Paul, M. J., & Dredze, M. (2013). Separating fact from fear: Tracking flu infections on twitter. In *Conference of the North American chapter of the association for computational linguistics: Human language technologies*, pp. 789−795.

Lampos, V. (2016). Flu detector: Estimating influenza-like illness rates from online user-generated content. arXiv preprint arXiv:1612.03494, 2016 - arxiv.org.

Lampos, V. (2017). Assessing public health interventions using Web content. arXiv:1712.08076. Available from <https://arxiv.org/pdf/1712.08076/> Accessed 20.06.18.

Lampos, V., Bie, T. D., & Cristianini, N. (2010). *Flu detector-tracking epidemics on Twitter*, *ECML PKDD 2010. Lecture Notes in Computer Science* (6323). Berlin, Heidelberg: Springer. Available from https://doi.org/10.1007/978-3-642-15939-8_42.

Lampos, V., Yom-Tov, E., Pebody, R., & Cox, I. J. (2015). Assessing the impact of a health intervention via user-generated Internet content. *Data Mining and Knowledge Discovery*, *29*, 1434. Available from https://doi.org/10.1007/s10618-015-0427-9.

Lane, I., Bryce, A., Ingle, S. M., & Hay, A. D. (2018). Does locally relevant, real-time infection epidemiological data improve clinician management and antimicrobial prescribing in primary care? A systematic review. *Family Practice*. Available from https://doi.org/10.1093/fampra/cmy008.

Laurent, M. R., & Vickers, T. J. (2009). Seeking health information online: Does Wikipedia matter? *Journal of the American Medical Informatics Association*, *16*(4), 471−479. Available from https://doi.org/10.1197/jamia.M3059.

Lee, B., Yoon, J., Kim, S., & Hwang, B.-Y. (2012). Detecting social signals of flu symptoms. In *Proceedings of the 8th IEEE international conference on collaborative computing: Networking, applications and worksharing*, pp. 544−545, IEEE. Print ISBN: 978-1-4673-2740-4.

Lee, K., Agrawal, A., & Choudhary, A. (2017). Forecasting influenza levels using real-time social media streams. *2017 IEEE international conference on healthcare informatics (ICHI)*, Park City, UT, pp. 409−414. Available from https://doi.org/10.1109/ICHI.2017.68.

Li, J., & Cardie, C. (2013). Early stage influenza detection from twitter. *Computer Science − Social and Information Networks, Computer Science − Computation and Language*. arXiv1309.7340L.

Li, T., Ding, F., Sun, Q., Zhang, Y., & Kinney, P. L. (2016). Heat stroke internet searches can be a new heatwave health warning surveillance indicator. *Scientific Reports*, *6*. Available from https://doi.org/10.1038/srep37294, Article number: 37294.

Li, Y. E., Tung, C.-Y., & Chang, S.-H. (2016). The wisdom of crowds in action: Forecasting epidemic diseases with a web-based prediction market system. *International Journal of Medical Informatics*, *92*, 35−43. Available from https://doi.org/10.1016/j.ijmedinf.2016.04.014.

Li, Z., Liu, T., Zhu, G., Lin, H., Zhang, Y., He, J., . . . Ma, W. (2017). Dengue Baidu Search Index data can improve the prediction of local dengue epidemic: A case study in Guangzhou, China. *PLoS Neglected Tropical Diseases*, *11*(3), e0005354. Available from https://doi.org/10.1371/journal.pntd.0005354.

Liu, K., Wang, T., Yang, Z., Huang, X., Milinovich, G. J., Lu, Y., . . . Lu, J. (2016). Using Baidu search index to predict Dengue outbreak in China. *Scientific Reports*, *6*. Available from https://doi.org/10.1038/srep38040, Article number: 38040.

Lu, F. S., Hou, S., Baltrusaitis, K., Shah, M., Leskovec, J., Sosic, R., . . . Santillana, M. (2018). Accurate influenza monitoring and forecasting using novel Internet data streams: A case study in the Boston Metropolis. *JMIR Public Health and Surveillance*, *4*(1), e4. Available from https://doi.org/10.2196/publichealth.8950.

Luo, Y., Zeng, D., Cao, Z., Zheng, X., Wang, Y., Wang, Q., & Zhao, H. (2010). Using multi-source web data for epidemic surveillance: A case study of the 2009 influenza A (H1N1) pandemic. In *Beijingin 2010 IEEE international conference on service operations and logistics and informatics*, Qingdao, China. Piscataway, NJ: IEEE. Available from https://doi.org/10.1109/SOLI.2010.5551614.

Madoff, L. C., Fisman, D. N., & Kass-Hout, T. (2011). A new approach to monitoring dengue activity. *PLoS Neglected Tropical Diseases*, *5*(5), e1215. Available from https://doi.org/10.1371/journal.pntd.0001215.

Madoff, L. C., & Li, A. (2014). Web-based surveillance systems for human, animal, and plant diseases. In R. M. Atlas, & S. Maloy (Eds.), *One health, people, animals and the environment*. Washington, DC: American Society for Microbiology. Available from https://doi.org/10.1128/microbiolspec.OH-0015-2012.

Malik, M. T., Gumel, A., Thompson, L. H., Strome, T., & Mahmud, S. M. (2011). Google flu trends and emergency department triage data predicted the 2009 pandemic H1N1 waves in Manitoba. *Canadian Journal of Public Health/Revue Canadienne de Santé Publique*, *102*(4), 294−297.

Marques-Toledo, Cd. A., Degener, C. M., Vinhal, L., Coelho, G., Meira, W., Codeço, C. T., & Teixei, M. M. (2017). Dengue prediction by the web: Tweets are a useful tool for estimating and forecasting dengue at country and city level. *PLoS Neglected Tropical Diseases*, *11*(7), e0005729. Available from https://doi.org/10.1371/journal.pntd.0005729.

Marquetoux, N., Stevenson, M. A., Wilson, P., Ridler, A., & Heue, C. (2016). Using social network analysis to inform disease control interventions. *Preventive Veterinary Medicine*, *126*, 94−104. Available from https://doi.org/10.1016/j.prevetmed.2016.01.022.

Martin, A., & Rino, S. M. (2016). Predicting the spread of pandemic influenza based on air traffic data and social media. *Norwegian University of Science and Technology, NTNU Open*. Available from <https://brage.bibsys.no/xmlui/bitstream/handle/11250/2415312/14623_FULLTEXT.pdf?sequence = 1/> Accessed 20.06.18.

Martin, L. J., Lee, B. E., & Yasui, Y. (2016). Google Flu Trends in Canada: A comparison of digital disease surveillance data with physician consultations and respiratory virus surveillance data. *Epidemiology and Infection*, *144*(2), 325−332. Available from https://doi.org/10.1017/S0950268815001478, Epub 2015 Jul 2.

Martino, S. D., Romano, S., Bertolotto, M., Kanhabua, N., Mazzeo, A., & Nejdl, W. (2017). Towards exploiting social networks for detecting epidemic outbreaks. *Journal of Flexible Systems Management*, *18*, 61. Available from https://doi.org/10.1007/s40171-016-0148-y.

Mavragani, A., & Ochoa, G. (2018). Forecasting AIDS prevalence in the United States using online search traffic data. *Journal of Big Data*, *5*, 17. Available from https://doi.org/10.1186/s40537-018-0126-7.

McClellan, C., Mir, A., Mutter, R., Kroutil, L., & Landwehr, J. (2017). Using social media to monitor mental health discussions—evidence from Twitter. *Journal of the American Medical Informatics Association*, *24*(3), 496−502. Available from https://doi.org/10.1093/jamia/ocw133.

McGough, S. F., Brownstein, J. S., Hawkins, J. B., & Santillana, M. (2017). Forecasting Zika incidence in the 2016 Latin America outbreak combining traditional disease surveillance with search, social media, and news report data. *PLoS Neglected Tropical Diseases*, *11*(1), e0005295. Available from https://doi.org/10.1371/journal.pntd.0005295.

McIver, D. J., & Brownstein, J. S. (2014). Wikipedia usage estimates prevalence of Influenza-like illness in the United States in near real-time. *PLoS Computational Biology*, *10*(4), e1003581. Available from https://doi.org/10.1371/journal.pcbi.1003581.

McMillan, N., Feng, J., Stamps, K., & Burr, R. E. (2014). Use of web-based symptom checker data to predict incidence of a disease or disorder. US Patent App. 14/180,683, US20140236613A1, US Application. Available from <https://patents.google.com/patent/US20140236613/> Accessed 20.06.18.

Merkord, C. L., Liu, Y., Mihretie, A., Gebrehiwot, T., Awoke, W., Bayabil, E., . . . Wimberly, M. C. (2017). Integrating malaria surveillance with climate data for outbreak detection and forecasting: The EPIDEMIA system. *Malaria Journal*, *16*, 89. Available from https://doi.org/10.1186/s12936-017-1735-.

Milinovich, G. J., Williams, G. M., & Hu, W. (2014). Internet-based surveillance systems for monitoring emerging infectious diseases. *The Lancet Infectious Diseases*, *14*(2), 160−168. Available from https://doi.org/10.1016/S1473-3099(13)70244-5.

Missier, P., Romanovsky, A., Miu, U., Pal, A., Daniilakis, M., Garcia, A., . . . Da Silva So, L. (2016). Tracking dengue epidemics using twitter content classification and topic modelling. In S. Casteleyn, P. Dolog, & C. Pautasso (Eds.), *Current trends in web engineering. ICWE 2016. Lecture notes in computer science* (Vol. 9881). Cham: Springer, arXiv:1605.00968v1 [cs.SI] 3 May 2016.

Moran, K. R., Fairchild, G. C., Generous, N., Hickmann, K., Osthus, D., Priedhorsky, R., . . . Del Valle, S. Y. (2016). Epidemic forecasting is messier than weather forecasting: The role of human behavior and internet data streams in epidemic forecast. *The Journal of Infectious Diseases*, *214*(suppl_4), S404−S408. Available from https://doi.org/10.1093/infdis/jiw375.

Morita M., Maskawa S., Aramaki E. (2013), Comparing social media and search activity as social sensors for the detection of influenza, In *The 5th international symposium on languages in biology and medicine (LBM 2013)*, Tokyo, Japan, December.

Moss, R., Zarebski, A., Dawson, P., & McCaw, J. M. (2016). Forecasting influenza outbreak dynamics in Melbourne from Internet search query surveillance data. *Influenza and Other Respiratory Viruses*, *10*(4). Available from https://doi.org/10.1111/irv.12376.

Moulton, A. D., & Schramm, P. J. (2017). Climate change and public health surveillance: Toward a comprehensive strategy. *Journal of Public Health Management & Practice*, *23*(6), 618−626. Available from https://doi.org/10.1097/PHH.0000000000000550.

Mowery, J. (2016). Twitter influenza surveillance: Quantifying seasonal misdiagnosis patterns and their impact on surveillance estimates. *Online Journal of Public Health Informatics*, *8*(3), e198. Available from https://doi.org/10.5210/ojphi.v8i3.7011.

Nagar, R., Yuan, Q., Freifeld, C. C., Santillana, M., Nojima, A., Chunara, R., & Brownstein, J. S. (2014). A case study of the New York City 2012-2013 Influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. *Journal of Medical Internet Research*, *16*(10), e236. Available from https://doi.org/10.2196/jmir.3416. , PMID: 25331122, PMCID: 4259880.

Nagel, A. C., Tsou, M.-H., Spitzberg, B. H., An, L., Gawron, J. M., Gupta, D. K., . . . Sawyer, M. H. (2013). The complex relationship of real space events and messages in cyberspace: Case study of Influenza and pertussis using tweets. *Journal of Medical Internet Research*, *15*(10), e237. Available from https://doi.org/10.2196/jmir.2705.

Nawa, N., Kogaki, S., Takahashi, K., Ishida, H., Baden, H., Katsuragi, S., . . . Ozono, K. (2016). Analysis of public concerns about Influenza vaccinations by mining a massive online question dataset in Japan. *Vaccine*, *34*(27), 3207−3213. Available from https://doi.org/10.1016/j.vaccine.2016.01.008.

Nawaz, M. S., Mustafa, R. U., & Lali, M. I. U. (2018). Role of online data from search engine and social media in healthcare informatics. In *Applying Big Data analytics in bioinformatics and medicine*, Chapter 11, pp. 272−293, IGI Global, 2017.

Neill, D. B., & Soetebier, K. A. (2011). Monitoring Twitter content related to influenza-like-illness in Spanish-speaking populations. *Emerging Health Threats Journal*, *4*(88), 88. Available from https://doi.org/10.3402/ehtj.v4i0.11185.

Nsoesie, E. O., & Brownstein, J. S. (2015). Computational approaches to influenza surveillance: Beyond timeliness. *Cell Host and Mocrob*, *17*(3), 275−278. Available from https://doi.org/10.1016/j.chom.2015.02.004.

Ocampo, A. J., Chunara, R., & Brownstein, J. S. (2013). Using search queries for malaria surveillance, Thailand. *Malaria Journal*, *12*, 390. Available from https://doi.org/10.1186/1475-2875-12-390.

Ofoghi, B., Mann, M., & Verspoor, K. (2016). Towards early discovery of salient health threats: A social media emotion classification technique. In *Proceedings of the Pacific symposium, biocomputing 2016*, pp. 504−515.

Ofran, Y., Paltiel, O., Pelleg, D., Rowe, J. M., & Yom-Tov, E. (2012). Patterns of information-seeking for cancer on the internet: An analysis of real world data. *PLoS ONE*, *7*(9), e45921. Available from https://doi.org/10.1371/journal.pone.0045921.

Olson, D. R., Konty, K. J., Paladini, M., Viboud, C., & Simonsen, L. (2013). Reassessing Google Flu Trends data for detection of seasonal and pandemic Influenza: A comparative epidemiological study at three geographic scales. *PLoS Computational Biology*, *9*(10), e1003256. Available from https://doi.org/10.1371/journal.pcbi.1003256.

Oren, E., Frere, J., Yom-Tov, E., & Yom-Tov, E. (2018). Respiratory syncytial virus tracking using internet search engine data. *BMC Public Health*, *18*, 445. Available from https://doi.org/10.1186/s12889-018-5367-z.

O'Shea, J. (2017). Digital disease detection: A systematic review of event-based internet biosurveillance systems. *International Journal of Medical Informatics*, *101*, 15−22. Available from https://doi.org/10.1016/j.ijmedinf.2017.01.019.

Parrella, A., Dalton, C. B., Pearce, R., Litt, J. C. B., & Stocks, N. (2009). ASPREN surveillance system for Influenza-like illness: A comparison with flutracking and the national notifiable diseases surveillance system. *Australian Family Physician*, *38*(11), 932−936.

Patwardhan, A., & Bilkovski, R. (2012). Comparison: Flu prescription sales data from a retail pharmacy in the US with Google Flu trends and US ILINet (CDC) data as flu activity indicator. *PLoS ONE*, *7*(8), e43611. Available from https://doi.org/10.1371/journal.pone.0043611.

Paul, M. J., & Dredze, M. (2011). You are what you Tweet: Analyzing Twitter for public health. In *Proceedings of the 5th international AAAI conference on weblogs and social media (ICWSM)*. Available from <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2880/3264/> Accessed 20.06.18.

Paul, M. J., & Dredze, M. (2012). *A model for mining public health topics from Twitter. Technical report*. Johns Hopkins University. Available from <https://pdfs.semanticscholar.org/41cb/ae26fe87307e6878e87b0a08056206a5c4c1.pdf/> Accessed 15.06.18.

Paul, M. J., Dredze, M., & Broniatowsk, D. A. (2014). Twitter improves Influenza forecasting. *PLOS Currents Outbreaks*. Edition 1. Available from https://doi.org/10.1371/currents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117.

Paul, M. J., Dredze, M., Broniatowski, D. A., & Nicholas, G. (2015). Worldwide influenza surveillance through twitter. In *AAAI workshop: WWW and public health intelligence*. Available from <https://pdfs.semanticscholar.org/6327/7acf07927625df96e668b8e812e6781f2a6b.pdf/> Accessed 20.06.18.

Paul, M. J., Sarker, A., Brownstein, J. S., Nikfajiam, A., Scotch, M., Smith, K. L., & Gonzalez, G. (2016). Social media mining for public health monitoring and surveillance. *Pacific symposium on biocomputing*, *Vol. 21*, 468.

Peek, N., Holmes, J. H., & Sun, J. (2014). Technical challenges for big data in biomedicine and health: Data sources, infrastructure, and analytics. *Yearbook of Medical Informatics*, *9*(1), 42−47. Available from https://doi.org/10.15265/IY-2014-0018. Published online 2014 Aug 15. PMCID: PMC4287098, PMID: 25123720.

Perrotta, D., Bella, A., Rizzo, C., & Paolotti, D. (2017). Participatory online surveillance as a supplementary tool to sentinel doctors for influenza-like illness surveillance in Italy. *PLoS ONE*, *12*(1), e0169801. Available from https://doi.org/10.1371/journal.pone.0169801.

Perrotta, D., Tizzoni, M., & Paolotti, D. (2017). Using participatory Web-based surveillance data to improve seasonal influenza forecasting in Italy. In *Proceedings of the 26th international conference on world wide web* (*WWW17*), pp. 303−310, Perth, Australia—April 03−07, 2017, International World Wide Web Conferences Steering Committee Republic and Canton of Geneva, Switzerland ©2017 table of contents. ISBN: 978-1-4503-4913-0. Available from https://doi.org/10.1145/3038912.3052670.

Petersen, J., Simons, H., Patel, D., & Freedman, J. (2017). Early detection of perceived risk among users of a UK travel health website compared with internet search activity and media coverage during the 2015−2016 Zika virus outbreak: An observational study. *BMJ Open*, *7*, e015831. Available from https://doi.org/10.1136/bmjopen-2017-015831.

Polgreen, P. M., Chen, Y., Pennock, D. M., Nelson, F. D., & Weinstein, R. A. (2008). Using internet searches for Influenza surveillance. *Clinical Infectious Diseases*, *47*(11), 1443−1448. Available from https://doi.org/10.1086/593098.

Pollett, S., Althouse, B. M., Forshey, B., Rutherford, G. W., & Jarman, R. G. (2017). Internet-based biosurveillance methods for vector-borne diseases: Are they novel public health tools or just novelties? *PLoS Neglected Tropical Diseases*, *11*(11), e0005871. Available from https://doi.org/10.1371/journal.pntd.0005871.

Pollett, S., Boscardin, W. J., Azziz-Baumgartner, E., Tinoco, Y. O., Soto, G., Romero, C., . . . Rutherford, G. W. (2017). Evaluating Google Flu Trends in Latin America: Important lessons for the next phase of digital disease detection. *Clinical Infectious Diseases*, *64*(1), 34−41. Available from https://doi.org/10.1093/cid/ciw657.

Prakash, B. A. (2016). Prediction using propagation: From flu trends to cybersecurity. *IEEE Intelligent Systems*, *31*(1), 84−88. Available from https://doi.org/10.1109/MIS.2016.1.

Priedhorsky, R., Osthus, D., Daughton, R., Moran, K. R., Generous, N., Fairchild, G., . . . DelValle, S. Y. (2017). Measuring global disease with Wikipedia: Success, failure, and a research agenda. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing* (*CSCW'17*), pp. 1812−1834, Portland, Oregon, USA—February 25−March 01, 2017, ACM New York, NY, USA ©2017, table of contents. ISBN: 978-1-4503-4335-0. Available from https://doi.org/10.1145/2998181.2998183.

Qiu, R., Hadzikadic, M., & Yao, L. (2017). *Estimating disease burden using google trends and wikipedia data advances in artificial intelligence: From theory to practice*, IEA/AIE 2017. *Lecture notes in computer science* (10351). Cham: Springer. Available from https://doi.org/10.1007/978-3-319-60045-1_39, Print ISBN 978-3-319-60044-439, Online ISBN 978-3-319-60045-1.

Radzikowski, J., Stefanidis, A., Jacobsen, K. H., Croitoru, A., Crooks, A., & Delamater, P. L. (2016). The measles vaccination narrative in twitter: A quantitative analysis. *JMIR Public Health and Surveillance*, *2*(1), e1. Available from https://doi.org/10.2196/publichealth.5059. Published online 2016 Jan 4. PMCID: PMC4869226, PMID: 27227144.

Rattanaumpawan, P., Boonyasiri, A., Vong, S., & Thamlikitkul, V. (2018). Systematic review of electronic surveillance of infectious diseases with emphasis on antimicrobial resistance surveillance in resource-limited settings. *American Journal of Infection Control*, *46*(2), 139−146. Available from https://doi.org/10.1016/j.ajic.2017.08.006.

Rekatsinas, T., Ghosh, S., Mekaru, S. R., Nsoesie, E. O., Brownstein, J. S., Getoor, L., & Ramakrishnan, N. (2017). Forecasting rare disease outbreaks from open source indicators. *The ASA Data Science Journal*, *10*(2), 136−150. Available from https://doi.org/10.1002/sam.11337.

Richterich, A. (2016). Using transactional big data for epidemiological surveillance: Google flu trends and ethical implications of 'infodemiology'. In B. Mittelstadt, & L. Floridi (Eds.), *The ethics of biomedical Big*

Data. Law, governance and technology series (29). Cham: Springer. Available from https://doi.org/10.1007/978-3-319-33525-4_3.

Robertson, C., & Yee, L. (2016). Avian influenza risk surveillance in North America with online media. *PLoS ONE*, *11*(11), e0165688. Available from https://doi.org/10.1371/journal.pone.0165688.

Robin, M.-H., Bancal, M.-O., Cellier, V., Délos, M., Felix, I., Launay, M., . . . Aubertot, J.-N. (2017). IPSIM-Web, an online resource for promoting qualitative aggregative hierarchical network models to predict plant disease risk: Application to brown rust on wheat. *APS Jourals, Plant Disease*, *102*(3), 488−499. Available from https://doi.org/10.1094/PDIS-12-16-1816-SR, March 2018.

Rohart, F., Milinovich, G. J., Avril, S. M. R., Cao, K.-A. L., Tong, S., & Hu, W. (2016). Disease surveillance based on Internet-based linear models: An Australian case study of previously unmodeled infection diseases. *Scientific Reports*, *6*. Available from https://doi.org/10.1038/srep38522, Article number: 38522.

Romano, S., Di Martino, S., Kanhabua, N., Mazzeo, A., & Nejdl, W. (2016). Challenges in detecting epidemic outbreaks from social networks. In *2016 30th international conference on advanced information networking and applications workshops (WAINA)*, NSPEC Accession Number: 16022255, Available from https://doi.org/10.1109/WAINA.2016.111

Sadilek, A., Kautz, H., DiPrete, L., Labus, B., Portman, E., Teitel, J., & Silenzio, V. (2016). Deploying nEmesis: Preventing foodborne illness by data mining social media. In *The IAAI conference on artificial intelligence (IAAI)*, pp. 3982−3990.

Sadilek, A., Kautz, H., & Silenzio V. (2012), Modeling spread of disease from social interactions. In *Proceedings of the ICSWM'11*. Available from <https://pdfs.semanticscholar.org/699c/3faeac25ebf5ebbb4e2641f3f4fb5a9a7720.pdf/> Accessed 20.06.18.

Sadilek, A., Kautz, H., & Silenzio, V. (2012). Predicting disease transmission from geo-tagged micro-blog data. In *Proceedings of the 26th AAAI conference on artificial intelligence*, December.

Samaras, L., García-Barriocanal, E., & Sicilia, M.-A. (2012). Syndromic surveillance models using Web data: The case of scarlet fever in the UK. *Informatics for Health and Social Care*, *37*(2), 106−124.

Samaras, L., García-Barriocanal, E., & Sicilia, M.-A. (2017). Syndromic surveillance models using web data: The case of influenza in Greece and Italy using google trends. *JMIR Public Health and Surveillance*, *3*(4), e90. Available from https://doi.org/10.2196/publichealth.8015.

Sandhu, R., Harsuminder, K. G., & Sandeep, K. S. (2016). Smart monitoring and controlling of Pandemic Influenza A (H1N1) using Social Network Analysis and cloud computing. *Journal of Computational Science*, *12*, 11−22. Available from https://doi.org/0.1016/j.jocs.2015.11.001.

Santillana, M., Nguyen, A. T., Dredze, M., Paul, M. J., Nsoesie, E. O., & Brownstein, J. S. (2015). Combining search, social media, and traditional data sources to improve Influenza surveillance. *PLoS Computational Biology*, *11*(10), e1004513. Available from https://doi.org/10.1371/journal.pcbi.1004513.

Santillana, M., Nguyen, A. T., Louie, T., Zink, A., Gray, J., Sung, I., & Brownstein, J. S. (2016). Cloud-based electronic health records for real-time, region-specific influenza surveillance. *Scientific Reports*, *6*. Available from https://doi.org/10.1038/srep25732, Article number: 25732.

Santillana, M., Nsoesie, E. O., Mekaru, S. R., Scales, D., & Brownstein, J. S. (2014). Using clinicians' search query data to monitor Influenza epidemics. *Clinical Infectious Diseases*, *59*(10), 1446−1450. Available from https://doi.org/10.1093/cid/ciu647.

Santos, J. C., & Matos, S. (2013). Predicting flu incidence from Portuguese Tweets. In *IWBBIO proceedings*, Granada, Spain, March 18−20, pp. 11−18.

Santos, J. C., & Matos, S. (2014). Analysing Twitter and web queries for flu trend prediction. *Theoretical Biology and Medical Modelling*, *11*(Suppl 1), S6. Available from https://doi.org/10.1186/1742-4682-11-S1-S6.

Schein, R., Wilson, K., & Kealan, J. (2011). *Literature review on effectiveness of the use of social media: A report for Peel Public Health*. [Region of Peel], Peel Public Health, 2011, Carleton University, Ottawa, Ontario, Canada.

Schellpfeffer, N., Collins, A., Brousseau, D. C., Martin, E. T., & Hashikawa, A. (2014). Web-based surveillance of illness in childcare centers. *Health Security*, *15*(5). Available from https://doi.org/10.1089/hs.2016.0124.

Schirmer, P., Lucero, S., Oda, G., Lopez, J., & Holodniy, M. (2010). Effective detection of the 2009 H1N1 Influenza pandemic in US Veterans Affairs medical centers using a national electronic biosurveillance system. *PLoS ONE*, *5*(3), e9533. Available from https://doi.org/10.1371/journal.pone.0009533.

Schwab-Reese, L., Hovdestad, W., Tonmyr, L., & Fluke, J. (2018). The potential use of social media and other internet-related data and communications for child maltreatment surveillance and epidemiological research: Scoping review and recommendations. *Child Abuse & Neglect*. Available from https://doi.org/10.1016/j.chiabu.2018.01.014, ISSN: 1873-7757, Publication Year: 2018, PMID:29366596.

Sciascia, S., & Radin, M. (2017). What can google and wikipedia can tell us about a disease? Big Data trends analysis in systemic lupus erythematosus. *International Journal of Medical Informatics*, *107*, 65−69. Available from https://doi.org/10.1016/j.ijmedinf.2017.09.002, November 2017.

Seifter, A., Schwarzwalder, A., Geis, K., & Aucott, J. (2010). The utility of "Google Trends" for epidemiological research: Lyme disease as an example. *Geospatial Health*, *4*, 135−137. Available from https://doi.org/10.4081/gh.2010.195.

Şerban, O., Thapen, N., Maginnisa, B., Hankina, C., & Foot, V. (2018). Real-time processing of social media with SENTINEL: A syndromic surveillance system incorporating deep learning for health classification. *Information Processing & Management*. Available from https://doi.org/10.1016/j.ipm.2018.04.011, ISSN: 0306-4573, Publication Year: 2018.

Seyyed Hosseini, S. (2018). An infodemiology study on breast cancer in Iran: Health information supply versus health information demand in PubMed and Google Trends. *Emerald Inside*, *36*(2), 258−269. Available from https://doi.org/10.1108/EL-03-2017-0062, 2018/04/03, N1 -doi:10.1108/EL-03-2017-0062.

Sharpe, J. D., Hopkins, R. S., Cook, R. L., & Strile, C. W. (2016). Evaluating Google, Twitter, and Wikipedia as tools for influenza surveillance using Bayesian change point analysis: A comparative analysis. *JMIR Public Health and Surveillance*, *2*(2), e161. Published online 2016 Oct 20. Available from DOI: 10.2196/publichealth.5901, PMCID: PMC5095368, PMID: 27765731.

Shikha, V., Younghee, P., & Mihui, K. (2017). Predicting flu-rate using big data analytics based on social data and weather conditions. *Advanced Science Letters*, *23*(12), 12775−12779. Available from https://doi.org/10.1166/asl.2017.10897. (5), Publisher: American Scientific Publishers.

Shin, S.-Y., Kim, T., Seo, D.-W., Sohn, C. H., Kim, S.-H., Ryoo, S. M., . . . Lim, K. S. (2016). Correlation between national influenza surveillance data and search queries from mobile devices and desktops in South Korea. *PLoS ONE*, *11*(7), e0158539. Available from https://doi.org/10.1371/journal.pone.0158539.

Signorini, A. (2015). *Use of social media to monitor and predict outbreaks and public opinion on health topics* (PhD thesis). University of Iowa. Available from <http://ir.uiowa.edu/etd/1503/> Asceesed 19.06.18.

Signorini, A., Segre, A. M., & Polgreen, P. M. (2011). The use of Twitter to track levels of disease activity and public concern in the US during the Influenza A H1N1 pandemic. *PLoS ONE*, *6*(5), e19467. Available from https://doi.org/10.1371/journal.pone.0019467.

Sivasankari, S., Kavitha, M., & Saranya, G. (2017). Medical analysis and visualisation of diseases using Tweet data. *Research Journal of Pharmacy and Technology*, *10*(12), 4306−4312. Available from https://doi.org/10.5958/0974-360X.2017.00788.0.

Sousa, L., De Mello, R., Cedrim, D., Garcia, A., Missier, P., Uchôa, A., . . . Romanovsky, A. (2018). VazaDengue: An information system for preventing and combating mosquito-borne diseases with social networks. *Information Systems*, *75*, 26−42.

Spruit, M., & Lytras, M. (2018). Applied data science in patient-centric healthcare: Adaptive analytic systems for empowering physicians and patients. *Telematics and Informatics*, *35*(4), 643−653. Available from https://doi.org/10.1016/j.tele.2018.04.002, 2018.

Suzumura, T. (2011). StreamWeb: Real-time web monitoring with stream computing. In *2011 IEEE international conference on web services*, July 4−9, INSPEC Accession Number: 12219322, Available from https://doi.org/10.1109/ICWS.2011.16.

Teng, Y., Bi, D., Xie, G., Jin, Y., Huang, Y., Lin, B., . . . Tong, Y. (2017). Dynamic forecasting of Zika epidemics using Google Trends. *PLoS ONE*, *12*(1), e0165085. Available from https://doi.org/10.1371/journal.pone.0165085.

Tilston, N. L., Eames, K. T. D., Paolotti, D., Ealden, T., & Edmunds, W. J. (2010). Internet-based surveillance of Influenza-like-illness in the UK during the 2009 H1N1 Influenza pandemic. *BMC Public Health*, *10*, 650. Available from https://doi.org/10.1186/1471-2458-10-650.

Tony, Yang Y., Horneffer, M., & DiLisio, N. (2013). Mining social media and web searches for disease detection. *Journal of Public Health Research*, *2*(1), 17−21. Available from https://doi.org/10.4081/jphr.2013.e4. Published online 2013 May 31. PMCID: PMC4140326, PMID: 25170475.

Valdivia, A., López-Alcalde, J., Vicente, M., Pichiule, M., Ruiz, M., & Ordobas, M. (2010). Monitoring Influenza activity in Europe with Google Flu Trends: Comparison with the findings of sentinel physician networks-results for 2009-10. *EuroSurveillance*, *15*(29). Available from https://doi.org/10.2807/ese.15.29.19621-en.

Verdery, A. M., Siripong, N., & Pence Brian, W. (2017). Social network clustering and the spread of hiv/aids among persons who inject drugs in 2 cities in the Philippines. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *76*(1), 26−32. Available from https://doi.org/10.1097/QAI.0000000000001485.

Visvizi, A., & Lytras, M. D. (Eds.), (2019). *Smart Cities: Issues and Challenges: Mapping Political, Social and Economic Risks and Threats*. Amsterdam, The Netherlands: Elsevier.

Walker, J. G. (2013). New media methods for syndromic surveillance and disease modelling. *CAB reviews perspectives in agriculture veterinary science nutrition and natural resources*, February. Available from https://doi.org/10.1079/PAVSNNR20138031.

Wang, F., Wang, H., Xu, K., Raymond, R., Chon, J., Fuller, S., & Debruyn, A. (2016). Regional level influenza study with geo-tagged Twitter data. *Journal of Medical Systems*, *40*, 189. Available from https://doi.org/10.1007/s10916-016-0545-y.

Wang, H.-W., Chen, D.-R., Yu, H.-W., & Chen, Y.-M. (2015). Forecasting the incidence of dementia and dementia-related outpatient visits with google trends: Evidence from Taiwan. *Journal of Medical Internet Research*, *17*(11), e264. Available from https://doi.org/10.2196/jmir.4516. Published online 2015 Nov 19. PMCID: PMC4704919, PMID: 26586281.

Wasim, A. (2018). *Using Twitter data to provide qualitative insights into pandemics and epidemics* (PhD thesis). University of Sheffield, UK, Available from <http://etheses.whiterose.ac.uk/20367/1/Final%20PhD%20Thesis%2011%20MAY.pdf/> Accessed 20.06.18.

Wilson, K., & Brownstein, J. S. (2009). Early detection of disease outbreaks using the Internet. *CMAJ*, *180*(8), 829−831. Available from https://doi.org/10.1503/cmaj.1090215.

Woo, H., Cho, H. S., Shim, E., & Lee, J. K. (2017). Identification of keywords from Twitter and web blog posts to detect influenza epidemics in Korea. *Disaster Medicine and Public Health Preparedness*. Available from https://doi.org/10.1017/dmp.2017.84, Published online: 31 July 2017.

Woo, H., Cho, Y., Shim, E., Lee, J.-K., Lee, C.-G., & Kim, S. H. (2016). Estimating influenza outbreaks using both search engine query data and social media data in South Korea. *Journal of Medical Internet Research*, *18*(7), e177. Available from https://doi.org/10.2196/jmir.4955.

Xiao, Q. Y., Liu, H. J., & Feldman, M. W. (2017). Tracking and predicting hand, foot, and mouth disease (HFMD) epidemics in China by Baidu queries. *Epidemiology & Infection*, *145*(8), 1699−1707. Available from https://doi.org/10.1017/S0950268817000231, June 2017.

Xie, Y., Chen, Z., Cheng, Y., Zhang, K., Agrawal, A., Liao, W.-K, & Choudhary, A. (2013). Detecting and tracking disease outbreaks by mining social media data. In *Proceedings of the 23rd international joint conference on artificial intelligence*. AAAI Press, pp. 2958−2960.

Xu, Q., Gel, J. R., Ramirez, L., Nezafati, K., Zhang, Q., & Tsu, K.-L. (2017). Forecasting influenza in Hong Kong with Google search queries and statistical model fusion. *PLoS ONE*, *12*(5), e0176690. Available from https://doi.org/10.1371/journal.pone.0176690.

Xu, W., Han, Z.-W., & Ma J. (2010). A neural network based approach to detect Influenza epidemics using search engine query data. In *e-Business engineering (ICEBE) 2011 IEEE 8th international conference*, pp. 9−15, Available from https://doi.org/10.1109/ICMLC.2010.5580851.

Yan, S. J., Chughtai, A. A., & Macintyre, C. R. (2017a). Effectiveness of web-based social sensing in health information dissemination—A review. *Telematics and Informatics*, *34*(1), 194−219. Available from https://doi.org/10.1016/j.tele.2016.04.012.

Yan, S. J., Chughtai, A. A., & Macintyre, C. R. (2017b). Utility and potential of rapid epidemic intelligence from internet-based sources. *International Journal of Infectious Diseases*, *63*, 77−87. Available from https://doi.org/10.1016/j.ijid.2017.07.020.

Yang, A. C., Huang, N. E., Peng, C.-K., & Tsai, S.-J. (2010). Do seasons have an influence on the incidence of depression? The use of an internet search engine query data as a proxy of human affect. *PLoS ONE*, *5*(10), e13728. Available from https://doi.org/10.1371/journal.pone.0013728.

Yang, M., Li, Y.-J., & Kiang, M. (2011). Uncovering social media data for public health surveillance, Association for Information Systems AIS Electronic Library (AISeL). In *PACIS 2011 Proceedings*. ISBN: [978-1-86435-644-1]; Full paper.

Yang, S., Kou, S. C., Lu, F., Brownstein, J. S., Brooke, N., & Santillana, M. (2017). Advances in using Internet searches to track dengue. *PLoS Computational Biology*, *13*(7), e1005607. Available from https://doi.org/10.1371/journal.pcbi.1005607.

Yang, S., Santillana, M., Brownstein, J. S., Gray, J., Richardson, S., & Kou, S. C. (2017). Using electronic health records and Internet search information for accurate influenza forecasting. *BMC Infectious Diseases*, *17*, 332. Available from https://doi.org/10.1186/s12879-017-2424-7.

Yang, W. (2017). *Early warning for infectious disease outbreak: Theory and practice* (p. 388) Cambridge, MA: Academic Press, ISBN: 978-0-12-812343-0.

Yanga, S., Santillana, M., & Kou, S. C. (2015). Accurate estimation of Influenza epidemics using Google search data via ARGO. *Proceedings of the National Academy Sciences*, *112*(47), 14473−14478. Available from https://doi.org/10.1073/pnas.1515373112, published ahead of print November 9, 2015.

Ye, X., Li, S., Yang, X., & Qin, C. (2016). Use of social media for the detection and analysis of infectious diseases in China. *ISPRS International Journal of Geo-Informatics*, *5*(9), 156. Available from https://doi.org/10.3390/ijgi5090156.

Yom-Tov, E., Borsa, D., Cox, I. J., & McKendry, R. A. (2014). Detecting disease outbreaks in mass gatherings using Internet data. *Journal of Medical Internet Research*, *16*(6), e154. Available from https://doi.org/10.2196/jmir.3156. PMID: 24943128, PMCID: 4090384.

Yom-Tov, E., Cox, I. J., & Lampos, V. (2015). Learning about health and medicine from Internet data. In *Proceedings of the 8th ACM international conference on web search and data mining (WSDM'15)*, pp. 417−418, Shanghai, China, February 02−06, ACM, New York, NY, ©2015, table of contents. ISBN: 978-1-4503-3317-7. https://doi.org/10.1145/2684822.2697042.

Yom-Tov, E., White, R. W., & Horvitz, E. (2014). Seeking insights about cycling mood disorders via anonymized search logs. *Journal of Medical Internet Research*, *16*(2), e65. Available from https://doi.org/10.2196/jmir.2664. PMID: 24568936, PMCID: 3961703.

Young, S. D., Rivers, C., & Lewis, B. (2014). Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. *Preventive Medicine*, *63*, 112−115. Available from https://doi.org/10.1016/j.ypmed.2014.01.024, ISSN: 1096-0260 Publication Year: 2014.

Young, S. D., Torrone, E. A., Urata, J., & Aral, S. O. (2018). Using search engine data as a tool to predict syphilis. *Epidemiology*, *29*(4), 574−578. Available from https://doi.org/10.1097/EDE.0000000000000836.

Yuan, Q., Nsoesie, E. O., Lv, B., Peng, G., Chunara, R., & Brownstein, J. S. (2013). Monitoring influenza epidemics in China with search query from Baidu. *PLoS ONE*, *8*(5), e64323. Available from https://doi.org/10.1371/journal.pone.0064323.

Zhang, Y., Bambrick, H., Mengersen, K., Tong, S., & Hu, W. (2018). Using Google Trends and ambient temperature to predict seasonal influenza outbreaks. *Environment International*, *117*, 284−291. Available from https://doi.org/10.1016/j.envint.2018.05.016.

Zhou, X., Yang, F., & Feng, Y. (2017). A spatial-temporal method to detect global influenza epidemics using heterogeneous data collected from the Internet. *Journal IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, *15*(3), 802−812. Available from https://doi.org/10.1109/TCBB.2017.2690631.

Zou, B., Lampos, V., & Cox, I. J. (2018). Multi-task learning improves disease models from web search. In *Proceedings of 2018 IW3C2 (International World Wide Web Conference Committee) (WWW'18)*, published under Creative Commons CCBY4.0 License. ACMISBN978-1-4503-5639-8/18/04, Available from https://doi.org/10.1145/3178876.3186050.

Zou, B., Lampos, V., Gorton, R., & Cox, I. J. (2016). On infectious intestinal disease surveillance using social media content. In *Proceedings of the 6th international conference on digital health conference (DH'16)*, pp. 157−161, Montréal, Québec, Canada, April 11−13. ACM New York, NY, ©2016, table of contents. ISBN: 978-1-4503-4224-7. Available from https://doi.org/10.1145/2896338.2896372.