International Journal of Public Health Research
2016; 4(3): 14-22
http://www.openscienceonline.com/journal/ijphr
ISSN: 2381-4829 (Print); ISSN: 2381-4837 (Online)

Open Science

Social Dialogue and Scientific Production on Big and Open Data in Health: From Facilitating Surveillance and Preventing Disease to Fostering Behavioral Changes

Marcelo D´Agostino[1, *], Felipe Mejía[2], Myrna Marti[3], David Novillo[4], Federico G. de Cosio[1], Nasim Farach[5]

[1]Health Information and Analysis Unit, Department of Communicable Diseases, Health Analysis, Pan American Health Organization (PAHO), Washington D. C., United States of America

[2]The Pan American Health Organization (PAHO), Bogotá, Colombia

[3]Department of Knowledge Management, Bioethics and Research, Pan American Health Organization (PAHO), Buenos Aires, Argentina

[4]Department of Knowledge Management, Bioethics and Research, Pan American Health Organization (PAHO), Washington D. C., United States of America

[5]International Public Health Consultant, Tegucigalpa, Honduras

Email address

dagostim@paho.org (M. D´Agostino)

[*]Corresponding author

To cite this article

Marcelo D´Agostino, Felipe Mejía, Myrna Marti, David Novillo, Gerardo de Cosio, Nasim Farach. Social Dialogue and Scientific Production on Big and Open Data in Health: From Facilitating Surveillance and Preventing Disease to Fostering Behavioral Changes. *International Journal of Public Health Research.* Vol. 4, No. 3, 2016, pp. 14-22.

Received: April 21, 2016; **Accepted**: April 28, 2016; **Published**: May 20, 2016

Abstract

Background: Big Data and Open Data are concepts that have evolved over the years and have different applications in health, from facilitating surveillance and preventing disease to fostering behavioral changes. *Objectives:* This exploratory study has two objectives: 1) To provide understanding about the use of Big Data and Open Data in the health arena and 2) to characterize the discussion, social behavior, interest and information shared in social networks and other online sources about Big Data and Open Data in health. *Methods:* Keywords related to "Health Big Data" and "Health Open Data" were used to gather data from social networks and from scientific databases. A qualitative based analysis was conducted upon a randomly selected subset of tweets, a list of comments in selected Facebook posts, a group of Instagram photographs and a list of selected LinkedIn posts. Trends in searches and publications were determined for Google Trends and the three scientific paper databases. *Results:* Majority of tweets are apps, the Cloud, consumers, costs, hackers, homecare, interoperability, mining, mobile, monitoring, openness, physicians, privacy, quality, research, safety, sensors, social media, startups, storage and wearables. Comments in selected Facebook posts showed mistrust toward any social network site involved in health. Users blamed companies for health security problems. On Instagram highly visual graphs and charts are the most common types of posts, followed by photographs of events related to data and health. LinkedIn posts including both Big Data and health count for more than 50,000 for "all time periods" and more than 3,000 for a period of "one week to one month". The most common topics for the first 20 results sorted by relevance were: partnerships, use of Big Data in health, health Big Data analytics, and Big Data trends. Google Trends showed that searches including health Big Data and health Open Data have increased steadily since 2012. For the three scientific paper databases, there seemed to be a greater increase in publications regarding Big Data than for Open Data. *Conclusions:* An integrated analysis of social tools and scientific databases can provide valuable insights into people's perceptions on innovative applications in health. Our results suggest that while interest in Big Data is increasing rapidly, there are growing concerns about data leakage and the ability of providers and governments to assure privacy. For Open Data in health, the increase in interest is not as large as that of Big Data in health.

Keywords

Big Data, Open Data, Open Governmental Data, Social Media, Qualitative Research

1. Introduction

Big Data is now considered a topic of great interest for its applications in public health and medicine, as well as in other arenas. The phrase "Big Data refers to large, diverse, complex, typically distributed data sets generated from various sources such as sensors, emails, social network websites, online transactions, and/or all other digital sources available today and in the future" [1]. However, over the years different definitions have been produced and there remains no universal agreement about what constitutes Big Data. [2] [3] Big Data implies three dimensions: volume, variety and velocity. [4] Volume is the magnitude of the data, variety is how heterogeneous a dataset might be and velocity is about how rapidly data is generated. [4] Other characteristics are variability, veracity, validity and volatility. [5] Positive evidence of Big Data use in health includes its use in preventing the onset of diseases, gathering information in order to detect risk factors and their possible intervention, and informing intervention designs for change in health-related behaviors. [6] Other possibilities include using Big Data for observational evidence, as a source for clinical decision support systems, customization of health services and deliverance customized health information to patients, as well as for epidemiological surveillance. [7] Social media is one of several sources of Big Data, and is generated through several platforms across the internet such as search engines (e.g. Google) and social networks (Facebook, Twitter and Instagram, among many others). The intersection of Big Data and social media has being integrated in one single concept: Social Big Data, [8] which has been defined as "Those processes and methods that are designed to provide sensitive and relevant knowledge to any user or company from social media data sources when data sources can be characterized by their different formats and contents, their very large size, and the online or streamed generation of information". [9]

Open Data primarily refers to granting free access to scientific information, and has evolved into a concept of non-copyright-restricted access to data delivered by governments. [10] It has also been defined as Public Sector Information. [11] The benefits of Open Data have been grouped into three categories: political and social, economic, and operational and technical. Some of the benefits within these categories are transparency and accountability, stimulation of competitiveness and innovation, and reuse and non-duplication of data. [12] Open Data in public health can improve surveillance by increasing non-governmental and external capacity analyses, as well as fostering community-based innovations in data analysis. [13]

As social Big Data emerges as a potential source of information that can be translated into knowledge for informing various interventions and strategies, there also exists the possibility of using data retrieved from social media for analyzing perceptions and attitudes toward a variety of topics.

In order to provide understanding about the possible impact of Big Data and/or Open Data in the health domain, an exploratory study was conducted over selected social networks and other online-based sources in order to characterize the discussions, social behavior, people's interest levels and information shared. This paper will present the results of this study in an attempt to capture the perception of users of social media internet channels.

2. Methods

2.1. Data Source

The sources used were Twitter, Facebook, Instagram, LinkedIn, Google Trends and scientific paper databases. The tool used for retrieving the data from Twitter between 6 October, 2015 and 3 November, 2015 was Tweet Binder, a web service that provides means to analyze, classify, report and share terms and hashtags on Twitter. For Instagram data retrieval, the tool used for retrieving the data from 1 through 30 October, 2015 was Mixagram, a web service that allows users to search more than one keyword at the same time. For the other sources of information, each website's own search engines were used. In the case of Facebook, the search was done between 11 and 12 November, 2015. Selected Top Posts, classified as such by Facebook, were further analyzed. In the case of LinkedIn, all results were retrieved, but only the first 20 ordered by relevance as given by the search engine were further analyzed. In the case of Google Trends and the scientific paper databases, all results were considered.

2.2. Keyword Search

For all sources provided above, the following search strategies were implemented: data AND health, datos AND salud, "Big Data"AND health, ""Big Data" AND salud, "Open Data" AND health, "datos abiertos" AND salud, "open government data" AND health, "datos gubernamentales abiertos" AND salud, and ("datos abiertos" AND gobierno AND salud).

2.3. Analysis

Twitter: For each search strategy during October, 2015, a quantitative approach was implemented to determine (a) the number of original tweets, (b) retweets, (c) replies, (d) number of links and pictures, (e) lists of the most active users based on those who sent the most tweets (including retweets), (f) those who retweeted the most, (g) those who tweeted original tweets the most, (h) a list of the most popular users based on those

who included the keywords and had the most followers, (i) a list of users with the highest impact based on tweets generating the highest impact (tweets x followers), (j) a list of the top 10 users based on those most mentioned and most retweeted, and (k) a list of the top 10 hashtags. For the first search strategy a qualitative approach was implemented. All the tweets from 6 October, 2015 through 3 November, 2015 were retrieved. A representative sample (with a confidence level of 99% and sample error of +/- 1%) was randomly selected from the population for further analysis. The randomization was conducted using the function: *random ()* in Microsoft Office Excel 2013. The sample was divided into groups. Each group was read and analyzed separately in order to minimize tedium. There were two groups in the morning and two groups in the afternoon throughout the three-day process. Notes were taken of key points, shared themes and common threads within each group.

Facebook: Facebook does not provide open information. In fact, results vary each time the same search is conducted. It is also important to point out that some results do not seem to directly relate to the search. In other words, given results sometimes do not seem to include the keywords used in the search phrase. In some cases, the keywords could be found inside the text, in the subject of the result, or in nearly any other location. Both a quantitative and qualitative approaches were implemented in each search strategy. The quantitative approach determined, characterized and organized the resulting number of groups, follower pages and latest posts. The qualitative approach selected top posts based on user engagement and those posts were further analyzed in search of key points, shared themes and common threads.

Instagram: A list of photographs from 1 through 31 October, 2015 were listed and classified both with quantitative and qualitative indicators such as number of likes and of comments and classifications according to the following categories: Dataviz, research/working group event, data device tracker, data app tracker, data tracking activity, data device/platform, data platform, data website and adds for telemedicine. 85 photographs uploaded to Instagram included the keywords data and health. However, these results also included hashtags that contained one of the keywords within them, such as #healthyiswealthy. These were not considered. After excluding those that did not apply to the overall theme, 36 photographs were considered for further analysis.

Linked In: Through LinkedIn's search engine, posts for each search strategy were retrieved, listed and classified based on date, relevance and topic type (applications, trends, uses, security, challenges, privacy, analytics, impacts, government, partnerships).

Google Trends: Using the Google Trends search engine, each search strategy was incrementally analyzed for the participation of any given term within the general concept. This analysis started with the keyword "Data" and how the keyword was related to Big Data and Open Data; from the results of Big Data and Open Data, the keyword's relation to Health Big Data and Health Open Data was analyzed; the same process concluded with Health Open Government Data.

Scientific Papers: A search of Virtual Health Library (VHL), PubMed and Science Direct was conducted. The number of scientific papers published in each available year were retrieved to check any incremental numbers of publications for the following topics of interest: Big Data, Open Data, open government data, health Big Data, health Open Data and health open government data. The search was conducted within summaries, titles and/or keywords.

3. Results

3.1. Twitter

More than 46,000 tweets in October, 2015 included the keywords "data and health." 35% of those tweets were retweets, and 48% included links and/or pictures. Of the more than 26,000 contributors, 81% tweeted only once. The most active accounts were related to hemp and cannabis-related processes. The most mentioned Twitter users, meaning when the user account is included within a tweet, were health policy-makers, health decision-makers and health institutions. The most common hashtags were associated to jobs, health care and mhealth.

More than 4,000 tweets in October, 2015 included the keywords "datos and salud." 45% of those were retweets and 47% included links and/or pictures. Among the more than 3,000 contributors, 90% tweeted only once. The most active accounts were related to news agencies. The most mentioned Twitter users were news agencies and one local international organization: PNDU Bolivia. The most common hashtags were related to animals, jobs and peace awareness.

More than 3,000 tweets in October, 2015 included the words Big Data and health. 38% of those were retweets and 57% included links and/or pictures. Of the more than 2,000 contributors, 86% tweeted only once. The most active accounts were specifically related to Big Data. The most mentioned Twitter users belonged to health organizations and news agencies. The most common hashtags were #BigData and #Health.

"Big Data and salud" and "Open Data and health" appeared in less than 300 tweets each in October, 2015, while "datos abiertos and salud" and "open government data and health" appeared in less than 10 tweets in the same month.

The topics most discussed in retrieved tweets are apps, the Cloud, companies (Apple, Google, Microsoft, Samsung), consumers, cost, hackers, homecare, interoperability, mining, mobile, monitoring, openness, physicians, privacy, quality, research, safety, sensors, social media, startups, storage and wearables. The tweets in Spanish expressed concern and complaints about data privacy. Apple and Microsoft were the most often mentioned, as were Cisco and IBM. Cloud services and apps were found in nearly all tweets.

Apps

Most tweets referred to the many health-related apps available to users, as well as their advantages in health management both for patients, caregivers and providers. Tweets expressing concerns about data protection and

complaints of loss of information were commonly found.

Cloud

The use of health-related cloud services is seen as both the next step and a challenge regarding privacy assurance. Some users complained about loss of information when using a cloud service for health management.

Consumers

Tweets express concern about the use of consumer health data, although not from the perspective of tech consumers. Some tweets promote the use and application of privacy guidelines.

Cost

Tweets mostly express the advantage of Big Data in reducing health care costs; others focus more on the risks, expressed as cost, of using Big Data analysis for health-related purposes.

Hackers

Hackers are seen as being particularly interested in health data. Also, some tweets address the issue of citizens worrying about hackers and health data.

Home

Several tweets focus on the impact of Big Data, mobile devices and other related technologies on home health care.

Interoperability

Interoperability is seen as one of the major challenges that has yet to be overcome, in spite of large investments by governments and health organizations.

Mining

Mining is a working concept used to express the possibilities of using data to improve health services. Some tweets focus on the challenges, others on the positive impacts, and some provide current examples of mining in use.

Mobile

Mobile devices and services integrated with data mining are expressed as keys to transforming and improving health care. Some tweets promote the use of mobile device health-related processes, while others express that users are willing to share their health data.

Monitor

Monitoring is also a working concept used to express how data, mobile devices and/or social media can be used to improve health care.

Open

Open is a working concept mostly used in the contexts of open source, open access and Open Data. Tweets express its use to improve health care, but others express its inability to do so.

Physicians

Physicians are the most commonly referred to personnel in tweets, many of which express the use and impact of Big Data and data mining on physicians' duties.

Privacy

Privacy of health data is easily the most mentioned in all tweets. There is general concern about how to secure health data, while at the same time using data mining to improve health care.

Quality

Tweets express the need for good data and/or data of quality in order to properly utilize Big Data and data-mining to improve health care.

Research

Tweets express various fields that research should focus on regarding health data, including Big Data data-mining. Some provide examples of current research.

Safety

Along with privacy, safety of health data is the most common topic. Concerns regarding companies' ability to secure health data are widespread.

Social Media

Social media is seen as key to data-mining for health care improvement. Twitter is regarded as the most commonly used social network for data mining, followed by Facebook.

Startup

Tweets on this topic are generally divided between those showing examples of startups using data-mining to challenge current health systems and those providing examples of startups failing due to health data challenges such as interoperability.

Storage

Storage is a topic of interest, specifically when it comes to electronic health records. It is also common to find lists of different ways for storing health data and tweets regarding security risks.

Wearable

"Wearable" is also a working concept sometimes used interchangeably with "mobile devices." Some tweets deal with wearables' use in monitoring health care and their integration with apps and data-mining, while others express concern about misuse of wearable devices regarding health data privacy.

Table 1. *Results in Twitter per search strategy.*

Type	data + health	datos + salud	Big Data + health	Big Data + salud	Open Data + health	datos abiertos + salud	open government data + health
Tweets	46.194	4.055	3.394	146	277	6	3
Text tweets	6.966	231	118	5	15	0	0
Replies	1.623	226	1.306	4	7	0	0
Retweets	16.001	1.842	50	48	217	3	0
Links & pictures	22.173	1.901	1.954	89	43	3	3
Impacts	246.011.289	133.177.004	14.013.368	261.369	1.424.660	12.028	1.919
Followers per contributor	5.412	11.385	4.431	1.912	4.029	2.005	640
Tweets per contributor	1,7	1,34	1,5	1,2	1,1	1	3

Type	data + health	datos + salud	Big Data + health	Big Data + salud	Open Data + health	datos abiertos + salud	open government data + health
Reach	146.969.682	34.347.919	10.253.186	237.142	1.039.373	12.028	1.919
Contributors	26.969	3.017	2.314	124	258	6	1
Number of tweets	Number of contributors						
1	21.868	2.724	2.001	109	247	6	3
2	3.016	172	180	9	7	0	0
3	888	41	49	5	1	0	0
4	405	20	22	1	2	0	0
5	218	18	9	0	1	0	0
6	154	7	15	0	0	0	0
+6	464	35	38	0	0	0	0

3.2. Facebook

Facebook is a social network that does not provide open information. In fact, results vary with each repetition of the same search. Some results do not seem be appropriately related to the search terms, meaning that a given result sometimes does not include the keywords used in the search. The keywords may be found inside the text of comments, in the subject of the results or in nearly any other position. The search was conducted only once on 11 November, 2015.

As with Twitter searches, the search phrase "data AND health" produced the most results. The two biggest Facebook groups have more than 6,000 members together. For the search Big Data and health, only two pages complied with the query. For "datos and salud," the only group found was more related to general information about fitness. This was expected, since as previously mentioned, datos is not a completely accurate translation for data. The only two pages found were about general health information.

The qualitative analysis shows mistrust toward any social network involved in health; users blame companies for health security problems, stating that the companies must have enough resources to keep health data secure. There is a perception that health data hackers come from within pharmaceutical companies wanting to use the data to sell their drug products.

First topic: Health Data Track.

- This post had 2,900 likes, 77 comments and 763 shares.
- Title: Tracking your own health data too closely can make you sick.

It is difficult to identify whether the position of those who commented is positive or negative. One person agrees and thinks that symptoms should not even be looked up online, but later states that this statement was a "funny comment." Another user states that "if it's on the internet, it has to be right," which can be understood as either a sarcastic position or a position disagreement with the topic's title. One user thought otherwise, stating, "The minute I started to pay attention to my health, I actually felt better" and another says, "obsessing over health can cause false symptoms."

Second topic: Social Network into Health.

- This post has 149 likes, 89 comments and 38 shares.
- Title: Is Facebook planning a move into health?

In general, comments express mistrust toward any social network involving itself in health. One user says, "I'm questioning the health of those who run Facebook!" or "no thanks. I don't wanna see a [sic] ad of doctors on my page [smiley emoticon]." Several users responded to the topic with, "never," "no never," "absurd," or "Facebook doesn't need my health data."

Third topic: Health Data Security

- This post has 198 likes, 33 comments and 223 shares.
- Title: Health insurer Anthem hacked; patient and employee data apparently exposed.

In general, users blame Anthem for the security problem, stating that Anthem must have enough resources to keep health data secure. One user states, "it's nice to know your internet security sucks as much as your coverage," and "with their rates, you would think they would have enough money to buy & [sic] staff a world class IT department." One user worries, "If they're careless with our personal information…how can they be trusted with our medical data."

Fourth topic: Health Data Hackers

- This post has 351 likes, 18 comments and 126 shares.
- Title: Why do hackers want your health data?

Although hackers are generally understood as bad, one user comments that, "not all hackers are criminals. The proper term is 'computer criminal'." One user states that these hackers come from pharmaceutical companies who want the data to sell drugs, stating, "hackers are really big pharma data mining for customers." Others point out countries are to blame for the hackers: "the Russian and chineese [sic] govt [sic] have thousand of hackers trying to break into every web site in America."

3.3. Instagram

From 1 through 31 October, 2015, 85 photographs that included both keywords were uploaded to this social network. However, these results also included hashtags that contain the keywords within them, such as #healthyiswealthy. These were not considered. After excluding those that did not apply to the inclusion criteria, 36 photographs were considered for further analysis. Dataviz, highly visual graphs or charts, are the most common posts, followed by photographs of events related to data and health. Tracking devices and tracking apps are

next in frequency, with wearable tracking devices being the most commonly posted. Others include tracking activities that do not require any type of technology, such as notes or pictures, followed by health data platforms and websites.

Table 2. *Classification of Instagram photographs during the period of study.*

Classification	#
Dataviz	9
Research/working group event	9
Data device tracker	7
Data app tracker	3
Data tracking activity	3
Data device/platform	2
Data platform	1
Data website	1
Telemedicine add	1

3.4. LinkedIn

For posts including "data and health," over 405,000 posts were found for "all time periods", as it is defined in the search engine filter. The most common topics for first the 20 results sorted by relevance were breach of data and security, ownership of data, and health and data challenges. More than 40,000 posts were found for the "one week to one month" period, as it is defined in the search engine filter. The most common topics for the first 20 results sorted by relevance were health data analytics, use of data for health services, and ownership of data.

For posts including "datos and salud," more than 5,000 posts were found for "all time periods". The most common topics for first 20 results sorted by relevance were use of data in health services, databases, and data privacy. Over 500 posts were found for the "one week to one month" period. The most common topics for the first 20 results sorted by relevance were Big Data in health and cybersecurity.

For posts including "Big Data and health," more than 50,000 posts were found for "all time periods". The most common topics for the first 20 results sorted by relevance were partnerships, use of Big Data in health, and health Big Data analytics. Over 3,000 posts were found for the "one week to one month" period. The most common topics for first 20 results sorted by relevance were health Big Data analytics and Big Data trends.

For posts including "Open Data and health," more than 2,000 posts were found for "all time periods". The most common topics for first the 20 results sorted by relevance were impacts of Open Data in health, Open Data and governments, and Open Data applications. Over 200 posts were found for the "one week to one month" period. The most common topic for the first 20 results sorted by relevance was applications of Open Data.

Other searches in Spanish represented less than 50 posts.

3.5. Google Trends

Big Data and Open Data returned far fewer results than searches using only Data. Prior to 2012, there were fewer queries for Big Data than for Open Data. Following that year, Big Data searches increased considerably over Open Data searches. Additionally, searches including "Health Big Data" and "Health Open Data" have increased steadily since 2012, especially those using "Health Big Data." Alternatively, "Open Government Data" searches have increased since 2010. Spanish results are similar to the searches conducted in English. "Datos Abiertos" searches remained stable during the period studied, and then experienced a drop in 2013.

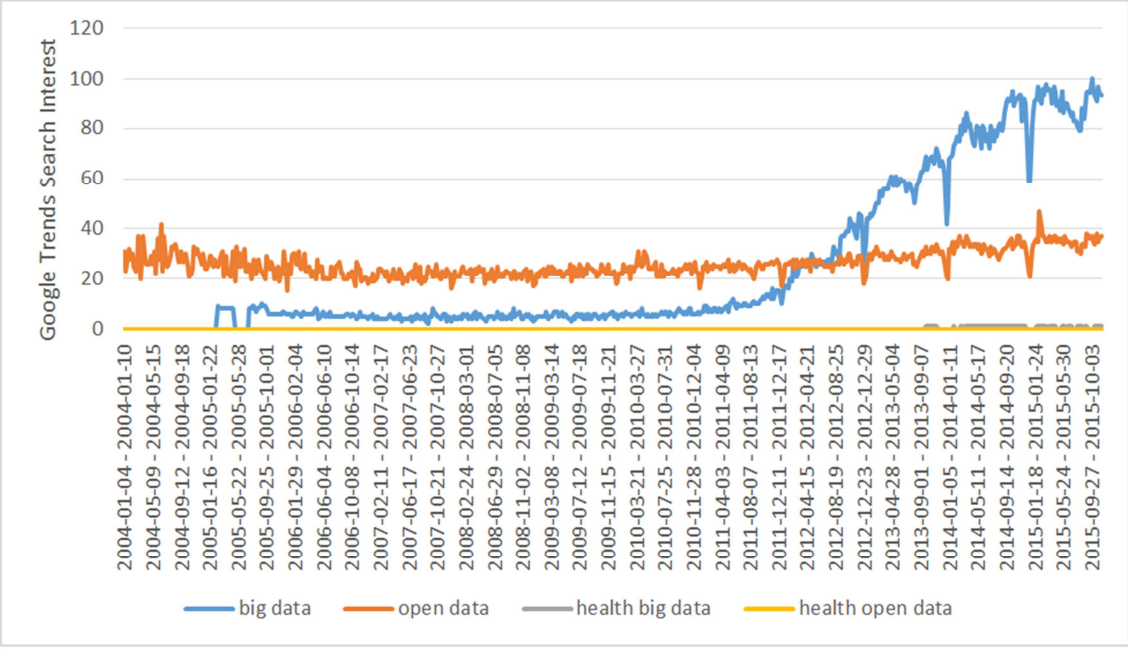

Fig. 1. *Searches of Big Data, Open Data, Health Big Data and Health Open Data based on Google Trends from January 2004 through September 2015.*

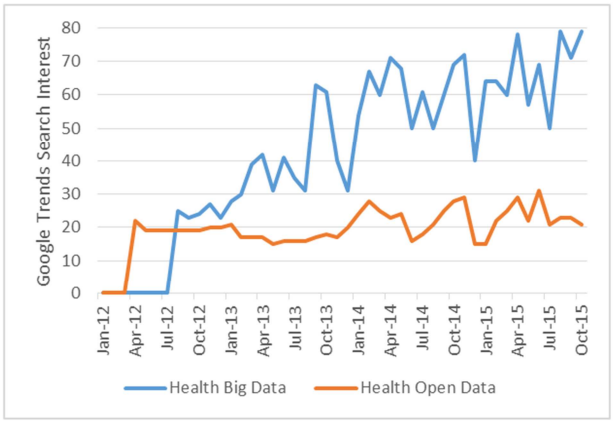

Fig. 2. Searches of Health Big Data and Health Open Data based on Google Trends from January 2004 through September 2015.

3.6. Scientific Papers

The number of publications in VHL increased from 20 publications in 2012 to 276 in 2015 for Big Data, and from 14 in 2012 to 46 in 2014 for Open Data. Only two publications were found dealing with Open Government. In Spanish, publications included four for Big Data and three for Open Data. Datos Masivos, Datos Abiertos y Datos Gubernamentales Abiertos had one, two and zero publications, respectively.

PubMed had 430 publications regarding Big Data in 2014; in the previous year, there were less than the half that number. For Open Data, there were 93 publications in 2014 and 46 in 2013. In Spanish, publications included five for Big Data and zero for Open Data. There were no publications including Datos Masivos, Datos Abiertos y Datos Gubernamentales Abiertos.

In Science Direct, there were 157 papers addressing Big Data in 2012, and 3,605 in 2015. "Big Data and Health" were represented in 46 publications in 2012 and 1,226 in 2015. This change is comparable to changes in Open Data papers: 199 in 2012, and 595 in 2015. "Open Data and Health" papers numbered 70 in 2012, and 231 in 2015. There were nine papers dealing with Open Government Data in 2012, and 28 in 2015.

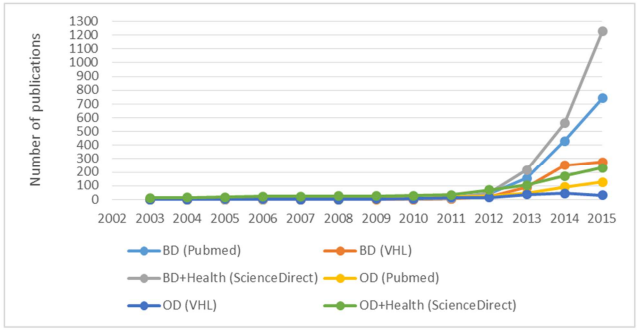

Fig. 3. Number of Publications per year in three databases including in their Abstracts, Titles and/or Keywords: Big Data and Open Data.

4. Discussion

In summary, we found that there has been a significant increase in interest about Big Data in general terms, and a growing interest in Big Data's uses within all aspects of health. The results also showed that interest in Open Data and Health Open Data has remained stable over the years. The period between 2011 and 2012 was the time during which both Big Data and Health Big Data overtook Open Data and Health Open Data, respectively. While publications about Open Data and Big Data have both increased over the years, the increase in Big Data was more accelerated. As expected based on the Open Data results, the interest in Open Government Data significantly increased in 2010, but has remained stable since then.

We also found that security and privacy are the topics of most interest and importance for users, being physicians the professions most related to these topics. There is a general perception that, while Big Data could reduce health care delivery costs, several security and privacy concerns remain that have not yet been resolved by tech providers; also, there is still not enough research to avoid data hacking and ensure quality control. On the other hand, there is a great degree of interest in apps for mobile phones, wearable devices, cloud services and sensor devices, which might imply a growing potential development area for public health. This is strongly related to finding that Big Data, as well as the devices and services, are perceived as being helpful to health self-management, particularly for fitness purposes. Retrieving information from social media databases has been getting more attention and interest from health care consumers. However, social media companies' activities regarding health are not positively perceived. Within the texts analyzed, the most- mentioned social media platform for health is Twitter, probably due to its Open Data availability. The most important challenges for Big Data and health are interoperability of information systems, and safe and secure data mining and storage.

Instagram, because of its image-based focus, offers information on user interests and needs, as well as mobile device use. Almost all images of health tracking devices were fitness-related, which also indicates a limited use of this network for broader public health approaches. However, the authors of this study consider that Instagram, when interconnected with other social networks such as Twitter, Facebook and LinkedIn, could be a very effective social communication tool for public health campaigns, due to the power of visual communication methods.

LinkedIn is more useful for understanding trends in research, currently popular topics and disseminating scientific knowledge in non-academic language. It gives professionals an opportunity to express opinions in a short essay format. This format was found in titles related to summaries of false health claims in Big Data, news of companies' partnerships, and current investigations by health and news agencies.

Google Trends helps users understand how a certain topic is growing in people's minds over the years and compare similar topics. The results obtained with social media, even when conducted only during the month of October, 2015 show that in the relationship of health Big Data and health Open Data, there is more growing interest and concern about the former than the latter.

Considerable increases in scientific publications may indicate the importance and relevance of Big Data and Open Data within the scientific community. The same cannot be said of the concept of Open Government Data.

Recommendations

A similar analysis during a longer period of study is recommended. Retrieving all tweets that include the keywords of interest, if possible, would enable researchers to foresee other trends and patterns not discernible within a retrieval period of only one month. It would also allow for more accurate comparisons with other means such as Google Trends. This recommendation applies to social media networks such as Instagram and LinkedIn, as well. In the case of Facebook, a more detailed qualitative analysis both in scope and depth, should be conducted in order to better understand public perception regarding Big Data and Open Data in the health arena. Other social networks could be considered, only if there is sufficient evidence that their users are interested in this topic of Big Data and Open Data as they relate to health.

Limitations

This study has several limitations. The period of study was too short in relation to the increasing search interest shown by Google Trends. The search strategy may have omitted important and meaningful tweets, posts and photographs. The high volume of tweets may be too much for a qualitative approach under the current paradigm, and may require new and innovative analysis methods in order to fully comprehend its meaning. Reading users' comments could potentially bias new comments from other users, in which case the comments, rather than being individual statements, are more likely to simply replicate previously posted ideas.

Twitter offers Open Data for several types of analysis, although it is more useful for quantitative than for qualitative analysis. The restriction of 140 characters per comment on Twitter might limit users' abilities to completely speak their minds. This is of particular note due to the fact that in almost all search strategies, more than 30% of the results were retweets. It is difficult to know whether a user retweeting another user's tweet completely understands the content of what is retweeted, or if the retweet is generated mechanically through any platform that can perform an automatic task.

Facebook data poses more challenges than Twitter in quantitative analysis due to its data access restrictions. However, it is preferable for qualitative analysis thanks to the possibility of long comments by users and long replies to comments. This represents something of a middle point between an open question from a questionnaire and a question from an in-depth interview.

5. Conclusion

Unstructured data from social networks, when combined with structured data and integrated with other tools such as Google Trends and scientific publication databases, might provide valuable insights about people's perceptions of new health applications and concepts. Our results suggest that while interest in Big Data is increasing rapidly, there are growing concerns about data leakage and the ability of providers and governments to assure privacy. Although interest in Open Data has not increased as much as that of Big Data in health, the surrounding concerns are similar. It is important for decision-makers to become more aware of the importance these growing concepts will have in public health. There is also a need to implement strategies to reduce public concerns about data privacy and security. Suggested areas of opportunity for public health involvement within these topics include data privacy, security of data, mobile application development and internet: social big data and health-related queries in search engines. People´s perception of an association among these concepts, combined with fitness and health self-management could be used as governmental strategies in the fight against non-communicable diseases (NCDs).

Future research should focus on using social media as a means to understand public perception and governmental adoption of both Big and Open Data in health, as well as for other qualitative studies.

References

[1] Hu, J., Yang, K., Warty, C., & Xu, K. (2015). Special issue on big data inspired data sensing, processing and networking technologies. Ad Hoc Networks, 35, 1–2. http://doi.org/10.1016/j.adhoc.2015.07.014

[2] Kacfah Emani, C., Cullot, N., & Nicolle, C. (2015). Understandable Big Data: A survey. Computer Science Review, 17, 70–81. http://doi.org/10.1016/j.cosrev.2015.05.002

[3] Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and Challenges of Big Data Research. Big Data Research, 2(2), 59–64. http://doi.org/10.1016/j.bdr.2015.01.006

[4] Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137–144. http://doi.org/10.1016/j.ijinfomgt.2014.10.007

[5] Asokan, G. V, & Asokan, V. (2015). Leveraging "big data" to enhance the effectiveness of "one health" in an era of health informatics. Journal of Epidemiology and Global Health, 311–314. http://doi.org/10.1016/j.jegh.2015.02.001

[6] Hansen, M. M., Miron-Shatz, T., Lau, a Y. S., & Paton, C. (2014). Big Data in Science and Healthcare: A Review of Recent Literature and Perspectives. Contribution of the IMIA Social Media Working Group. Yearbook of Medical Informatics, 9(1), 21–6. http://doi.org/10.15265/IY-2014-0004

[7] Luna, D. R., Mayan, J. C., García, M. J., Almerares, A. A., & Househ, M. (2014). Challenges and Potential Solutions for Big Data Implementations in Developing Countries. IMIA Yearbook of Medical Informatics, 36–41. http://doi.org/10.15265/IY-2014-0012

[8] Camacho, D., & Jung, J. J. (2016). Guest Editorial : Social big data with information fusion. Information Fusion, 28(August 2015), 44. http://doi.org/10.1016/j.inffus.2015.07.008

[9] Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2015). Social big data: Recent achievements and new challenges. Information Fusion, 28, 45–59. http://doi.org/10.1016/j.inffus.2015.08.005

[10] Kassen, M. (2013). A promising phenomenon of open data: A case study of the Chicago open data project. Government Information Quarterly, 30(4), 508–513. http://doi.org/10.1016/j.giq.2013.05.012

[11] Zuiderwijk, A., & Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. Government Information Quarterly, 31(1), 17–29. http://doi.org/10.1016/j.giq.2013.04.003

[12] Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. Information Systems Management, 29(4), 258–268.

[13] Turbelin, C., & Boëlle, P.-Y. (2013). Open data in public health surveillance systems: A case study using the French Sentinelles network. International Journal of Medical Informatics, 82(10), 1012–1021. http://doi.org/10.1016/j.ijmedinf.2013.06.009