



Leveraging “big data” to enhance the effectiveness of “one health” in an era of health informatics

Asokan G.V.^{a,*}, Vanitha Asokan^{b,1}

^a Public Health Program, College of Health Sciences, University of Bahrain, PO Box 32038, Bahrain

^b Pediatrics Department, American Mission Hospital, Manama, PO Box 1, Bahrain

Received 23 December 2014; received in revised form 1 February 2015; accepted 4 February 2015

Available online 5 March 2015

KEYWORDS

One health;
Big data;
Zoonoses;
Health informatics

Abstract Zoonoses constitute 61% of all known infectious diseases. The major obstacles to control zoonoses include insensitive systems and unreliable data. Intelligent handling of the cost effective big data can accomplish the goals of one health to detect disease trends, outbreaks, pathogens and causes of emergence in human and animals.

© 2015 Ministry of Health, Saudi Arabia. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The science of health *informatics deals with* how health information is captured, transmitted and utilized for healthcare delivery [1]. Data that is rapid, complete, reliable and abundant generates useful health information. Big data is the structured and unstructured data characterized by volume, complexity, diversity, and timeliness. Ever-growing big data exceeds the processing capacity of conventional database systems. In 2011, the total volume of healthcare data were estimated at 150 billion gigabytes, and an increase of 1.2–

2.4 billion gigabytes is expected annually [2,3]. The aim of this dispatch is to discuss the utility of big data to enhance the effectiveness of the systems approach-based one health.

2. Zoonoses and one health

Annually, 16 million human deaths occur worldwide due to infectious causes [4]. Zoonoses constitute 61% of all known infectious diseases and 75% of emerging diseases [5]. Jones et al. [6] reported that there have been 335 emerging disease events progressively increasing over the decades, mostly in developing regions of the world, and this pattern is expected to continue. In addition, an estimated 20% of all human illness and death occurred in the least developed countries attributed to

* Corresponding author. Tel.: +973 17435837.

E-mail addresses: agvaithinathan@uob.edu.bh (G.V. Asokan), kidzdoc311@gmail.com (V. Asokan).

¹ Tel.: +973 36878035.

endemic zoonoses [7]. The major obstacles in the prevention and control of zoonoses include:

- insensitive systems to detect outbreaks
- incomplete flow of reliable data
- poor inter-sectoral coordination
- inadequate infrastructure, use of technology and human resources

One health [8] is advocated to circumvent the obstacles through a systems approach-based collaborative effort of multiple disciplines working locally, nationally, and globally to attain optimal health for people, animals and the environment. A special need exists today to enhance the efficacy of one health by leveraging unexplored data sources for zoonoses that can rapidly and accurately transmit data and information to detect disease trends, outbreaks, pathogens and causes of emergence.

3. Big data

Capturing, analyzing, and sharing health data is difficult, expensive and often incoherent in a traditional system, whereas big data is transforming science. The availability of the vast amount of health data advances real-time tracking of diseases, predicting disease outbreaks, and developing healthcare. Big data networks work on a single task simultaneously by handling distributed resources with resiliency, consistency and application awareness to provide a robust evidence-informed health policy. The big data network is aiming to enhance the value by squeezing more actionable information [9]. More accurate analyses precede confident decision making, greater operational efficiencies, cost reductions and reduced risk.

Big data is a mix of unstructured and multi-structured data that comprises a large volume of information.

- i. *Unstructured data* are text heavy information that is not organized or easily interpreted by traditional databases or data models, e.g., metadata, twitter tweets on flu, and other social media posts.
- ii. *Multi-structured data* comes from a variety of data formats and types. It can also be derived from interactions between people and web applications or social networks, e.g., web log data, which is a combination of text and visual images along with structured data like form or transactional information [10].

The characteristics of big data [10,11] are:

- a. *Volume*: Transaction-based data stored over the years and unstructured data from social media contribute to the enormous volume of data. The

central issues are the relevance and using analytics to create value from data.

- b. *Velocity*: Spatial and temporal data streams reacting in real time to deal with volume and velocity are a challenge that can be overcome by RFID tags, sensors, smart metering and sampling of the data.
- c. *Variety*: Data sources are heterogeneous that come from structured numeric data, information from applications, unstructured text documents, PDFs, email, video, audio, mobile apps, etc. Storage, mining, cleansing, merging, linking, matching, transforming, analyzing and governing different varieties of data are an undertaking.
- d. *Variability*: Data flows can be consistent with regular peaks or inconsistent, such as daily, seasonal and event-triggered data trending in social media. Unstructured data and unexpected peaks are truly challenging.
- e. *Veracity*: Veracity refers to truthfulness that is devoid of the biases and abnormality. The strategy is to keep out 'dirty data' from building up in the systems.
- f. *Validity*: Validity refers to data correctness and accuracy for the intended use.
- g. *Volatility*: How long are the data valid and being stored? One needs to determine at what point is data obsolete to the current needs.

4. Big data and one health

One health promotes knowledge and accountability. Knowledge expands by the exchange of data and information, and accountability depends on transparency of information; lack of transparency leads to bias. Transparency of data requires synergy of systems and sharing across systems. Big data for one health that can synergize different systems encompass: Google (Maps, Trends, Translate); ProMED-mail; satellite images; application of geostatistics; habitat and migration of the birds, animals and marine species; agriculture; climate change; climate risk management; temperature; rainfall; floods; cyclones; land use; forest cover; recycling; air quality; waste management, etc.; and integration of WAHID-OIE [12], EMPRES-FAO [13] and WHO [14].

The emphasis on big data has increased. With the low cost of data generation, intelligent handling of data can contain more useful information. Big data offer new insights around networks, spatial and temporal dynamics of understanding human, animal and environmental systems at the systemic level, and for detecting interactions and nonlinearities among variables [15].

Despite the enormous benefits big data offers, one health scientists may be concerned to find relevant data and to understand the context of shared data in the data deluge. Further, accepting the new research tools to analyze large

pre-existing datasets rather than a hypothesis-driven prospective study will require a new mindset. Some of the common barriers encountered in harnessing big data include: difficulties in arranging the unstructured data, lack of technical expertise and skills in quantitative and statistical methods, data isolated within the institutions, lack of data privacy, challenges in high performance analytics (predictive analytics, data mining, text mining, forecasting and optimization), processing information without human checks, and data spiraling out of control may lead to invalid conclusions and unsound one health policies [16].

But the concerns and barriers can be circumvented by utilizing the tools to handle massive data sets and analytical techniques to generate hypotheses, indicate trends, identify anomalies, and make predictions. In one health epidemiology, the identification of ‘who infected whom’ allows us to quantify key characteristics, such as within and between species transmissions, transmission rates, incubation periods, the duration of infectiousness, and the high-risk groups. Interpretation of these data highlights the confluence of technology with mathematical and statistical approaches to enhance the understanding of transmission and control of zoonoses, for example, Google Flu Trends [17]; Project of disease containment in the Ivory Coast used Orange cellular phone data [18]; and Julia Salzman’s discovery of “circular RNAs” has the potential to understand and monitor diseases [19].

Recently, using the available online data, social media and local news reports, an algorithm developed by Health Map [20] indicated early signs of the Ebola disease spread in West Africa. Therefore, control of the current Ebola epidemic in West Africa requires utility of health ecosystem data and geomapped Ebola incidence data for a clinical decision support system [21,22], data treatment and information treatment to galvanize the flow of unstructured data (event-based surveillance) and structured data (indicator-based surveillance).

In summary, leveraging big data can accomplish the goals of one health rapidly with minimal cost for healthy people, animals and the environment.

Conflict of interest

None declared.

Acknowledgments

We thank the following from the College of Health Sciences, University of Bahrain: Dr. Aneesa Al-Sindi, Dean, for the encouragement and support provided

to work on this manuscript, and the additional support of Mrs. Raja AQ, Chairperson, Allied Health Department.

References

- [1] American Health Information Management Association. What is Health Information? <http://www.ahima.org/careers/healthinfo> [accessed Nov 15, 2014].
- [2] TechAmerica Foundation’s Federal Big Data Commission. Demystifying big data: a practical guide to transforming the business of government. Washington (DC): TechAmerica Foundation; 2012, <http://breakinggov.sites.breaking-media.com/wp-content/uploads/sites/4/2012/10/TechAmericaBigDataReport.pdf> [accessed Sep 25, 2014].
- [3] Hughes G. How big is “big data” in healthcare? A shot in the arm blog. SAS Health and Life Sciences blog Web site. <http://blogs.sas.com/content/hls/2011/10/21/how-big-is-big-data-in-healthcare/> [accessed Nov 2, 2014].
- [4] Editorial. Microbiology by numbers. *Nat Rev Microbiol* 2011;9:628.
- [5] Taylor LH, Latham SM, Woolhouse ME. Risk factors for human disease emergence. *Philos Trans R Soc Lond B Biol Sci* 2001;356:983–9.
- [6] Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. *Nature* 2008;451:990–3.
- [7] Grace D, Jones B, McKeever D, Pfeiffer D, Mutua F, Jemimah M, et al. Zoonoses: Wildlife/livestock interactions. Zoonoses (Project 1). A final report to the Department for International Development, UK and ILRI, Nairobi, Kenya by ILRI, Nairobi and Royal Veterinary College, London; 2011.
- [8] One Health. <http://www.avma.org/onehealth/responding.asp> [accessed Oct 18, 2014].
- [9] Heitmueller A, Henderson S, Warburton W, Elmagarmid A, Pentland AS, Darzi A. Developing public policy to advance the use of big data in health care. *Health Aff* 2014;33:1523–30.
- [10] Forbes. What is big data? <http://www.forbes.com/sites/lisaarthur/2013/08/15/what-is-big-data/> [accessed Sep 9, 2014].
- [11] Normandeau K. Beyond volume, variety and velocity is the issue of big data veracity. <http://inside-bigdata.com/author/kevin/> [accessed Oct 21, 2014].
- [12] World Organization for Animal Health. World Animal Health Information database interface. <http://www.oie.int/wahis/public> [accessed Aug 5, 2014].
- [13] Food and Agriculture Organization. Emergency prevention system. <http://www.fao.org/ag/againfo/programmes/en/empres/home.asp> [accessed Sep 2, 2014].
- [14] World Health Organization. Global atlas. <http://apps.who.int/globalatlas/DataQuery/default.asp> [accessed Aug 15, 2014].
- [15] Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. *Science* 2014;343:1203–5.
- [16] Hoffman S, Podgurski A. Big bad data: law, public health, and biomedical databases. *J Law Med Ethics* 2013; 41(Suppl. 1):56–60.
- [17] Goel S, Hofman JM, Lahaie S, Pennock DM, Watts DJ. Predicting consumer behavior with Web search. *Proc Natl Acad Sci U S A* 2010;107:17486–90.
- [18] Big data analytics. http://www.webopedia.com/TERM/B/big_data_analytics.html [accessed Aug 10, 2014].

- [19] Stanford conference focuses on big data in health care. <http://www.sfgate.com/technology/article/Stanford-conference-focuses-on-big-data-in-health-5499724.php> [accessed Sep 29, 2014].
- [20] Health Map. Ebola outbreaks; 2014. <http://healthmap.org/ebola/> [accessed Oct 11, 2014].
- [21] Weber GM, Mandl KD, Kohane IS. Finding the missing link for big biomedical data. *JAMA* 2014;311:2479–80.
- [22] Mandl KD. Ebola in the United States: EHRs as a public health tool at the point of care. *JAMA*. Published online October 20, 2014. <http://dx.doi.org/10.1001/jama.2014.15064>.

Available online at www.sciencedirect.com

ScienceDirect