Life Sciences, Society and Policy

Disease surveillance data sharing for public health: the next ethical frontiers

CrossMark

Patty Kostkova [ORCID]

Abstract

In the recent years, we have been witnessing a digital revolution in public and global health creating unprecedented opportunities for epidemic intelligence and public health emergencies. However, these opportunities created a double edge sword as access to data, quality monitoring and assurance, as well as governance and regulation frameworks for data privacy are lagging behind technological achievements.

In this paper we identify three ethical challenges: sharing data across various early warning tools to support risk assessment. Secondly, define the challenges to be addressed by the legal frameworks for public health data sharing to unlock the potential of population-level datasets for research with no impact on citizens privacy. The third challenge lies with stricter regulation of the IT industry with regards to manipulating user data - such an initiative, GDPR, comes to force in the EU in May 2018.

Introduction

In the recent years, we have been witnessing a digital revolution in public and global health. It is estimated there will be around 5.07 billion people worldwide using a mobile phone in 2019.[1] The use of mobile phones for accessing information about health has almost doubled since 2010 (when it was 17%) to 31% in 2013. According to Pew Research (Kostkova et al. 2014), this includes 52% of smartphone owners.

The rapid advancements in the digital health technologies and affordability of mobile phones have brought a dramatic shift in delivering public health interventions (Kostkova 2015), identifying and managing major outbreaks such as ebola (Vorovchenko et al. 2017), dengue (Albinati et al. 2017) and zika (McGough et al. 2017; Kostkova et al. 2010) and responding to healthcare emergencies. Web 2.0 technologies and real-time Big Data streamed and shared from social media, mobile phones and wearable/tracking devices have dramatically reshaped the delivery of healthcare, opportunities for managing personal health conditions and improving wellbeing. Specifically, for the international infectious disease surveillance landscape (Valente 2010) relying on the traditional epidemic intelligence systems Big Data

created an epistemic shift (WHO 2014). Real-time data have been successfully used for early warning systems (De Quincey and Kostkova 2010; Szomszor et al. 2010; Lampos and Cristiani 2012), the role of Twitter has been highlighted as a game changer (St Louis and Zorlu 2012), participatory surveillance crowdsourcing reporting to citizens at national and international levels (Guerrisi et al. 2016), emergency and risk communication (Kostkova et al. 2014) and have provided a challenging online space for public discourse of important health concerns such as vaccination (Salathé and Khandelwal 2011; Kostkova et al. 2016b; Kostkova et al. 2017).

However, these opportunities created a double edge sword as access to data, quality monitoring and assurance, as well as governance and regulation frameworks for data privacy are lagging behind technological achievements. Regulatory frameworks and evidence for the actual impact on public health and quantifiable improvement of health outcomes as a result of mHealth and big data remains limited. In this paper we outline three major ethical and governance challenges for digital epidemiology in the twenty-first century.

New ethical frontiers

Data ownerships and sharing for public health surveillance and relevant governance and lack of regulation of data have crystallized into three major ethical challenges.

Correspondence: p.kostkova@ucl.ac.uk
Institute for Risk and Disaster Reduction (IRDR), UCL, Gower Street, London WC1E 6BT, UK

IT systems successfully improved public health early warning using the new media but implementation of these systems in a real world surveillance by WHO, CDC, ECDC etc. keeps falling short. Sharing data across various early warning tools to support risk assessment and calibration of predictive models remains the **first ethical challenge**.

Further, two disparate citizens approaches emerged: government-regulated clinical and research, data is subject to high scrutiny by legal frameworks and sharing is often hampered by public mistrust. On the other hand, private user-generated health data collected from social media, apps, online searches and wearable devices seem having no shortage of users volunteering to share their lives in public domain or with IT and MedTech industry (Kostkova et al. 2016). Therefore **the second challenge** relates to defining the legal frameworks for public health data sharing to unlock the potential of these population-level datasets for research with no impact on citizens privacy. **The third challenge** lies with much stricter regulation of the IT industry with regards to manipulating user data - such an initiative, GDPR, comes to force in the EU in May 2018.

Surveillance data: Risk assessment and response saves lives

Real-time location-aware non-medical data sources (social media, personal traces, shopping lists, mobile data) could be mined for improving early warning systems, analyzed for threats detection to assist public health experts in risk assessment and response. From monitoring population mobility using mobile phones to fight human-transmitted infections (e.g. H1N1, Ebola), to tracking crowds during the Boston marathon bombing, or coordinating emergency aid delivery and humanitarian relief operations during the 2010 Haiti earthquake disaster (Maier 2014), new data are helping to transform early warning, coordination and response. Risk communication in case of healthcare emergencies and epidemics was also shown to benefit from the use of social media (Szomszor et al. 2011).

While there is unquestionably a potential in early-warning using Big data for early warning which has been the focus of most research endeavours by academics and industry (Barboza et al. 2013). We need a radically different integrated solution connecting independent systems through shared data and functionality, rather than continuing with the current isolated IT surveillance systems lacking interoperability and common data standards. Such novel tools leveraging the opportunities from data sharing for enable risk assessment and rapid response by frontline healthcare professionals are urgently required, as the recent Ebola outbreak demonstrated. Calibration of models in real time through combining different datasets has been outlined for the zika surveillance and early warning (Beltrán et al. 2018). Without successful integration of new Big Data systems with traditional epidemic intelligence and routine surveillance to aid risk assessment and response processes assessing the severity of outbreak and confidence in the predictions (Kostkova et al. 2014), the response will remain slow and disjoined (Moon et al. 2015).

Sharing data for public health

The **Big Data in healthcare** (including large linked data from electronic patient records as well as streams of real-time geo-located health data collected by personal wearable devices, etc.) and the **Open Data** (movement enabling sharing datasets) are creating new challenges around ownership of personal data while opening new research opportunities and drives for commercial exploitation (Kostkova 2015).

It has been highlighted that population level surveillance data sharing could enable faster and better coordinated response during health emergencies while opening new frontiers for data-driven research in public health (Kostkova 2013). However, enabling access to population level data, defining an internationally enforceable governance framework to benefit public health remains a major challenge despite national surveillance systems collecting data on notifieable diseases, and several international frameworks are making it compulsory to share such population-level data among states and with WHO.

Firstly, full transparency and clarity of public health data sharing requires active public engagement and better understanding of benefits and risks of data sharing (as defined by the Fundamental requirement for DPA Principle 1), strong transparent disclosure, and notification mechanism informing public about potential violations. Enforcement of these principles in the current legislative and regulatory framework remains a challenge (Kostkova et al. 2016b). These challenges have been outlined at the national level by the UK House of Parliament briefing highlighting the challenges between UK and EU legislation, governance conflicts of interest and open issues around data privacy and security (Houses of Parliament, Parliamentary Office of Science and Technology, Big Data and Public Health 2014).

At international level, it has been identified how multiple barriers for data sharing include technical challenges, motivational and economic issues, and political, legal and ethical considerations - each of these requires a spectrum of actions to be agreed and addressed (Panhuis et al. 2014).

At the EU level, the core EU legislation (Commission Decision 2008), and national level equivalents (Roush et al. 1999) constitutes of the Early Warning and Response System (EWRS), defined by EC decision 2000/57/EC and amended by decision 2008/351/EC, defining steps

required to be taken by EU member states (MS) in case of health threats of international importance. While surveillance and epidemic intelligence at European level were improved, the EWRS information nor data from ECDC-run dedicated surveillance networks (DSN) are shared with other MS and professional communities due to the MS control of national level data in the networks and confidential nature of the EWRS.

At international level, recent Chatham House study on data sharing for public health emergencies highlighted the disparity of public health data created by sub-regional and regional surveillance networks that should be integrated into any global framework, and the limitations of the WHO established International Health Regulations (IHR). Taking into account the social, political and cultural context of data sharing for public health with transparency and trust will be fundamental for any future success (Edelstein and Sane 2015).

Striking a balance between data sharing, personal data protection, stakeholder needs, and public good in order to ensure an effective global health response in real-time emergency situations remains a key challenge (Kostkova et al. 2016a).

IT and MedTech industry: The big brother
Digital traces are increasingly becoming essential signal sources for public health surveillance that add value by providing additional information. They include search keywords, loyalty purchase cards, sensor networks, drugs purchases, and mobile phone data (Kostkova et al. 2014).

Although no longer in operation, the Google Flu Trend project was one of the first to demonstrate a potential of online searches at Google search engine for tracking flu by comparing the signal to ILI surveillance network (ILINet) (Cook et al. 2011). However, this search data remains proprietary and is therefore not available for reproducility of the results, validation of results essential for transparent research, nor for the development of non-commercial applications. Google's acquired London-based AI start-up - DeepMind - got caught in a controversial legal battle for the illegal use of EPR data from a collaboration with Royal Free. In a widely medialized case, the UK Information Commissioner (ICO) declared the use of patient data illegal.[2]

The use of non-commercial search data has been piloted by the National electronic Library of Infection (Madle et al. 2006) demonstrating correlation of professional information needs, expressed through online searches, with policy changes around major outbreaks over several years (Kostkova et al. 2013).

However, the most alarming ethical and social concern is the fact that never has so much data about so many citizens been held by so few with little policy and legal

oversight, such as the IT industry, MedTech and mobile industries (Lupton 2014). As this paper goes to press, the IT giant Facebook is being rocked by the 'Cambridge Analytica' data sharing scandal dramatically shifting citizens' perceptions of online data sharing and driving the matter to governments' agendas (reference: Timothy B. Lee - 3/20/2018, Facebook's Cambridge Analytica scandal, explained Arstechnica, online).[3] Further, a new international regulatory framework bringing a radical shift in the direction of regulation of data usage by industry giving control back to users generating the data will be introduced with the EC legislation, GDPR,[4] coming to force on 25th May 2018. The GDPR provides the following rights for individuals: adding over DPA the right to erasure, the right to restrict processing, the right to data portability and the rights in relation to automated decision making and profiling making the use of user data for commercial purposes much harder.

Recently, there has been an interest in developing public data sharing platforms offering 'safe' data storage for customers driven by users' **changing attitudes** towards data privacy and increasing lack of support for data sharing and usage such as the 'Midata coop' initiative (Hafen et al. 2014) empowering users to share their data in strictly regulated data warehouses (von Grätz and Hafen 2016), and calls for system oversight (Vayena 2018). However, with the ncreasing influence of the big IT businesses on governments these steps would have ground-breaking industry implications (Kostkova et al. 2016b).

Conclusion
We have highlighted the potential of big data for public health bringing a epistemic shift for routine surveillance, early warning and response. We outlined the three major ethical and governance challenges preventing the full implementation in day to day public health services at national and international levels. These include: lack of support for risk assessment and response operations, data sharing and governance challenges, and data privacy and proprietary use by the IT and MedTech industrial players. As GDPR goes to force in the EU, more international attention ought to be given to regulatory and ethical aspects of Big Data to take the full advantage of data for public health for the benefit of citizens globally.

Endnotes
[1]https://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide/.
[2]https://www.computing.co.uk/ctg/news/3013129/google-deepmind-patient-data-sharing-deal-with-royal-free-nhs-trust-deal-was-illegal-rules-ico?utm_medium=email&utm_campaign=CTG.Daily_RL.EU.A.U&utm_source=

CTG.DCM.Editors_Updates&im_edp=ucl.ac.uk&im_ company=.

[3]https://arstechnica.com/tech-policy/2018/03/face-books-cambridge-analytica-scandal-explained/.

[4]https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/.

Author's contribution
PK, the author, has conducted this research exercise and written up the manuscript. The author read and approved the final manuscript.

Author's information
PK is a currently a Senior Lecturer in UCL Institute for Risk and Disaster Reduction (IRDR), UCL London, conducting research into digital epidemiology and the use of mobile serious games for community engagement in Brazil, Nigeria and Nepal.

Ethics approval and consent to participate
Not applicable

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

References
Albinati J, Meira W, Jr., Pappa GL, Teixeira M, Marques-Toledo C. Enhancement of Epidemiological Models for Dengue Fever Based on Twitter Data, Proceedings of DH '17 Proceedings of the 2017 International Conference on Digital Health, pp. 109–118, New York: ACM; 2017. https://dl.acm.org/citation.cfm?id=3079464.

Barboza P, Vaillant L, Mawudeku A, Nelson NP, Hartley DM, Madoff LC, Linge JP, Collier N, Brownstein JS, Yangarber R, Astagneu P. Evaluation of epidemic intelligence systems integrated in the earl alert in and reporting project of the detection of A/H5N1 influenza events. PLoS One. 2013;8(3):e57272.

Beltrán JD, Boscor A, dos Santos WP, Massoni T, Kostkova P. ZIKA: A new system to empower health workers and local communities to improve surveillance protocols by e-learning and to forecast Zika virus in real time in Brazil, In the proceedings of the 8th Digital Helath 2018, Lyon, pp. 90–94, https://doi.org/10.1145/3194658.3194683

Commission Decision 2008/426/EC. https://publications.europa.eu/en/publication-detail/-/publication/3e53de24-26d6-4645-b9ab-3931f3874c9e/language-en.

Cook S, Conrad C, Fowlkes AL, Mohebbi MH. Assessing google flu trends performance in the United States during the 2009 influenza virus a (h1n1) pandemic. PLoS One. 2011;6:8.

De Quincey E, Kostkova P. Early warning and outbreak detection using social networking web-sites: The potential of twitter. In Proceedings of the 2nd International ICST Conference on Electronic Healthcare, P. Kostkova Ed., Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, vol. 27, Berlin, Heidelberg: Springer; 2010. pp. 21–4. https://link.springer.com/chapter/10.1007/978-3-642-11745-9_4.

Edelstein M, J Sane. Overcoming Barriers to Data Sharing in Public Health: A Global Perspective - Project Centre on Global Health Security, Strengthening Data Sharing for Public Health. April 2015, Online at: https://www.chathamhouse.org/publication/overcoming-barriers-data-sharing-public-health-global-perspective#sthash.AKIDfQ9M.dpuf

Guerrisi C, Turbelin C, Blanchon T, Hanslik T, Bonmarin I, Levy-Bruhl D, Perrotta D, Paolotti D, Smallenburg R, Koppeschaar C, Franco AO, Mexia R, John Edmunds W, Sile B, Pebody R, van Straten E, Meloni S, Moreno Y, Duggan J, Kjelsø C, Colizza V. Participatory syndromic surveillance of influenza in Europe. J Infect Dis. 2016;214(Supplement 4):S386–92, Oxford University Press.

Hafen E, Kossmann D, Brand A. Health data cooperatives - citizen empowerment. Methods Inf Med. 2014;53(2):82–6. https://doi.org/10.3414/ME13-02-0051.

Houses of Parliament, Parliamentary Office of Science and Technology, Big Data and Public Health. POSTnote 474 July 2014 (accessed: researchbriefings.files.parliament.uk/documents/POST-PN-474/POST-PN-474.pdf. Accessed 2017.

Kostkova P. A Roadmap to Integrated Digital Public Health Surveillance: the Vision and the Challenges In the Proceedings of the 1st International Workshop on Public Health in the Digital Age: Social Media, Crowdsourcing and Participatory Systems (PHDA 2013), held in conjunction with 22nd WWW 2013, Rio de Janeiro, Brazil, May 2013, pp. 687–693

Kostkova P. Public Health. A chapter in Twitter: A Digital Socioscope. Eds. Yelena Mejova, Ingmar Weber, Michael W Macy. Cambridge: Cambridge Press; 2015. pp. 111–30. https://www.researchgate.net/publication/275658492_Public_Health.

Kostkova P. Grand Challenges in Digital Health, Front Public Health. 2015;3:134. Published online 2015 May 5. doi: https://doi.org/10.3389/fpubh.2015.00134

Kostkova P, Brewer H, de Lusignan S, Fottrell E, Goldacre B, Hart G, Koczan P, Knight P, Marsolier C, McKendry RA, Ross E, Sasse A, Sullivan R, Chaytor S, Stevenson O, Velho R, Tooke J. Who Owns the Data? Open Data for Healthcare. *Front Public Health*. 2016a;4:7. https://doi.org/10.3389/fpubh.2016.00007.

Kostkova P, De Quincey E, Jawaheer G. The potential of social networks for early warning nad outbreak detection systems: the swine flu Twitter study. Int J Infect Dis. 2010;14:e384–5, Elsevier.

Kostkova P, Garbin S, Moser J, Pan W. Integration and visualization public health dashboard: the medi+board pilot project. WWW (Companion Volume); 2014. pp. 657–62.

Kostkova P, Fowler D, Wiseman S, Weinberg JR. Major infection events over 5 years: How is media coverage influencing online information needs of health care professionals and the public? J. Med Internet Res. 2013;15:7.

Kostkova P, Mano V, Larson HJ, Schulz WS. VAC Medi+ board: Analysing Vaccine Rumours in News and Social Media, Proceedings of the 6th International Conference on digital Health, New York: ACM Press; 2016b. pp. 163–4. https://dl.acm.org/citation.cfm?id=2896338.2896370.

Kostkova P, Mano V, Larson HJ, Schulz WS. Who is Spreading Rumours about Vaccines?: Influential User Impact Modelling in Social Networks. In: Proceedings of the 2017 International Conference on Digital Health; 2017. pp. 48–52. https://doi.org/10.1145/3079452.3079505.

Kostkova P, Szomszor, M., and St. Louis, C. 2014. #swineflu: The use of twitter as an early warning and risk communication tool in the 2009 swine flu pandemic. ACM Trans Manage Inf Syst. 5, 2, Article 8 (July 2014), 25 pages. DOI: https://doi.org/10.1145/2597892

Lampos V, Cristiani N. Nowcasting events from the social web with statistical learning. ACM Trans Intell Syst Technol. 2012;3:4.

Lupton D. Health promotion in the digital era: a critical commentary. Health Promot Int (2014) doi: https://doi.org/10.1093/heapro/dau091 First published online: October 15, 2014

Madle G, Kostkova P, Mani-Saada J, A Roy - Lessons learned from evaluation of the use of the National electronic Library of Infection, Health Informatics J, Vol 12, No 2, pp. 137–151, Sage 2006

Maier P. Crisis Mapping Haiti: Some Final Reflections. April 2014, available online https://www.ushahidi.com/blog/2010/04/14/crisis-mapping-haiti-some-final-reflections

McGough SF, Brownstein JS, Hawkins JB, Santillana M. Forecasting Zika Incidence in the 2016 Latin America Outbreak Combining Traditional Disease Surveillance with Search, Social Media, and News Report Data. PLoS Negl Trop Dis. 2017;11(1):e0005295. Available at: http://bit.ly/2ixatVK

Moon S, et al. Will Ebola change the game? Ten essential reforms before the next pandemic. The report of the Harvard-LSHTM Independent Panel on the Global Response to Ebol. Lancet. 2015;386(Issue 10009):2204–21. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(15)00946-0/fulltext.

Panhuis AN, et al. A systematic review of barriers to data sharing in public health. BMC Public Health. 2014;14:1144.

Roush S, Birkhead G, Koo D, et al. Mandatory reporting of diseases and conditions by health care professionals and laboratories. JAMA. 1999;282:164–70.

Salathé M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. PLoS Comput Biol. 2011;7(10):e1002199.

St Louis C, Zorlu G. Can Twitter predict disease outbreaks? BMJ (Online). 2012;344. https://www.bmj.com/content/344/bmj.e2353.

Szomszor M, Kostkova P, De Quincey E. #Swineflu: Twitter predicts swine flu outbreak in 2009. In Proceedings of the 3rd International Conference on

Electronic Healthcare, M. Szomszor and P. Kostkova, Eds. Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, vol. 69. Berlin: Springer; 2010. pp. 18-26.

Szomszor M, Kostkova P, St Louis C. Twitter informatics: Tracking and understanding public reaction during the 2009 swine flu pandemics. In: Proceedings of the IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology (WI-IAT'11), vol. 1; 2011. pp. 320–3.

Valente TW. Social Networks and Health. Oxford: Oxford University Press; 2010.

Vayena E, Blasimme A. Health Research with Big Data: Time for Systemic Oversight. The Journal of Law, Medicine & Ethics. 2018;46(1):119–29. https://doi.org/10.1177/1073110518766026

von Grätz PG, Hafen E. 'A move towards freedom from digital feudalism' Insights HIMSS Europe, the IT Journal for Healthcare Leaders, June 2016. https://www.himssinsights.eu/move-towards-freedom-digital-feudalism

Vorovchenko T, Ariana P, van Loggerenberg F, Amirian P. #Ebola and Twitter. What Insights Can Global Health Draw from Social Media? In: Amirian P, Lang T, van Loggerenberg F, editors. Big Data in Healthcare. Cham: SpringerBriefs in Pharmaceutical Science & Drug Development. Springer; 2017.

WHO. 2014. Epidemic intelligence – Systemic event detection. http://www.who.int/csr/alertresponse/epidemicintelligence/en/index.html