Transformation of Big Data for Health in the European Union.

Andrew Boilson

andrew.boilson@dcu.ie

Anthony Staines

anthony.staines@dcu.ie

Justin Connolly

Justin.connolly@yahoo.co.uk

Paul Davis

paul.davis@dcu.ie

Regina Connolly

Regina.Connolly@dcu.ie

Topic Areas:

Healthcare Information Systems

Technologies to Promote a Healthy & Secure Society

Transformation of Big Data for Health in the European Union.

Abstract

Modern healthcare is increasingly dependent on good data, and effective information systems, for care delivery, and to develop and evaluate health policy. The context of big data differs in significant ways from traditional types of health data, while the use of big data for epidemiology and public health is becoming more common, the use of these tools for health service planning and health policy making lags behind. A large EU funded project (titled MIDAS) that focuses on merging, analysing and visualising data from heterogeneous sources to support health policy makers work in using and accessing health data across EU countries is underway. This paper briefly describes the key challenges that must be met to access, use and make sense of this big data in healthcare, focusing on legal, governance and ethical issues. Unless these issues are dealt with, the promise of Big Data for health, will never be fulfilled.

Keywords: Data Analytics, Epidemiology, Public Health

1.0 Introduction

Digital epidemiology is motivated by the same objectives as traditional epidemiology, but focuses on electronic data sources that emerged with the advent of information technology (Hay et al., 2013). It draws on developments such as the online growth of sharing platforms, which constantly generate vast amounts of data containing health related information (Vayena et al., 2015). Utilizing available global real-time data accelerated disease outbreak detection is feasible. Reports of emerging outbreaks have been detected through digital surveillance channels in advance of official reports e.g. Ebola virus outbreak, West Africa (Anema et al., 2014), assessment of health behaviour and attitudes (Salathe & Khandelwal 2011) and pharmacovigilance (White et al., 2013**)**.

Big Data technology and services are expected to grow worldwide at a compound annual growth rate of 40% (Vessat, 2012). The European Commission adopted the communication on the data-driven economy, focusing on the digital economy, innovation and services as drivers for growth and jobs and called for EU action to provide the right framework conditions for a single market for Big Data and cloud computing (Com, 2014, 442). They are also bringing in a new general EU legal framework on the protection of personal data. The General Data Protection Regulation aims at eliminating fragmentation and providing consistency and coherence for the whole Union (Com, 2012, 11). This is critical for the wider reuse of personal data in Europe.

The European Commission Directorate-general for Health and Consumers Protection (DG SANCO) are committed to capturing the potential of Big Data in public health policy and research to produce policy recommendations to member states according to the logic improvement of healthcare systems and in light of the Directive 2011/24/EU on Patients' Rights in Cross-Border Healthcare (Union, E.P.a.C.o.t.E., *DIRECTIVE 2011/24/EU on the application of patients' rights in cross-border healthcare*, 2011). In this context, they have funded the MIDAS project, which seeks to develop tools for accessing analysing and interpreting heterogeneous data sources, oriented towards health policy makers (Rankin et al. 2017).

2.0 Big Data and Public Health

Big data cannot be readily grouped into clearly demarcated functional categories. Depending on how they are queried and combined with other datasets, a given dataset can traverse categories in unpredictable ways (Vayena et al., 2015**)**. New data analytics constantly change the kinds of outcomes that become possible.

They go beyond early identification, and detection of disease patterns to include predictions of the event's trajectory or likelihood of recurrence (Thomas, 2014; Brockmann & Helbing, 2014**)**. These possibilities render good data governance, which is essential for ethical use. The last two decades have seen an explosion in big data throughout the healthcare value chains, as well as the advent of new platforms, tools, and methodologies in storing, structuring, and analysing big data (Bernstein, 2014).

Numerous questions can be addressed with big data analytics. Clinical outcomes may be predicted and/or estimated based on vast amounts of historical data, such as length of stay; patients who will choose elective surgery; patients at risk of surgical complications; risk for sepsis, MRSA, C. difficile, or other hospital acquired illness (Manyika et al., 2011).

Within public health applications included analysis of disease patterns and tracking disease outbreaks and transmission to improve surveillance and speed response (Raghupathi and Raghupathi, 2014). Faster development of more accurately targeted vaccines, e.g., choosing the annual influenza strains; and, turning large amounts of data into actionable information that can be used to identify needs, provide services, predict and prevent crises for the benefit of populations (Manyika et al., 2011). Optimisation of pharmaceutical outcomes utilising a variety of statistical techniques ranging from predictive modelling to machine learning, data mining is another area of rapid development. (Hernandez and Zhang, 2017).

The MIDAS project, data analytics and operations research applications will inform practitioners and policy makers solving healthcare problems. Supporting the development of sustainable long-term solutions across disease management, service delivery, and health policies, in part by optimizing the performance of system elements and analysing interactions (Capan et al., 2017)

3.0 Big Data Challenges

The big data dynamic environment generates challenges that relate not only to the value of health systems, individual rights and other moral requirements (Vayena et al., 2015). The distinct features are the methods by which data are generated, collected and stored. The kind of information that is inferred by their analysis, and eventually how that information is translated into practice (Neff, 2013). Data analytics technology can derive value in ways that were previously impossible, the technical capabilities have reached a level of sophistication and pervasiveness that demands careful consideration and presents several challenges.

3.1 Governance

Public health surveillance and public health research are governed by national and international legislation and guidelines. Many of these norms were developed in response to technologies that have now been superseded (Fairchild & Bayer, 2004). Such mechanisms may not be appropriate or effective in addressing the new ethical challenge posed nor the questions that will be raised if big data is effectively integrated into standard public health systems. Health research utilizing social media data and other online datasets have already exerted pressure on existing research governance procedures (Vayena et al., 2012).

As the amount of health related data and global digital information grows, so does the number of actors accessing and using this information. Assurances must be given that personal data related to health will be used appropriately, in the context of the intended uses and according to relevant laws. There is public scepticism about "where my data goes", "by whom it is used" and "for what purpose" in the fragmented and complex legal environment of the EU (EU Report, Use of Big Data in Public Health Policy Research, 2014).

A strong governance model and adaptation of best practices of new technologies are essential for the deployment of big data for large-scale data production, coupled with interoperable data storage, data integration, and data analytics solutions (Meldolesi et al., 2016). Further, evaluation of data needs to be assessed and interpreted in a timely manner (Hood et al., 2014; Raghupathi et al., 2014) to improve the efficiency, effectiveness, prediction and prevention strategies of public health services and policy (Rumsfeld et al., 2016; Monteithet al., 2016).

3.2 Confidentiality & Security

An important condition for the access to patient related information is the protection of personal data. Data protection rules as transposed into national laws in Europe do not yet establish full-harmonized conditions for health data processing, although this will improve as the General Data Protection Regulation (GDPR) is adopted (Maldoff, 2016).

Existing computational infrastructures can readily cope with the storage of big data, but the specific challenge for the EU is the lack of a suitable large-scale European infrastructure and methods of secure data distribution in a cross-border setting (Georgatos et al., 2013**)**.

It is crucial to ensure the infrastructures that exist and evolve are coordinated and sustainable. There are significant differences of culture and practice within and across Europe for data access and data, sharing polices (DLA Piper, 2016). Recently a code of practice on the secondary use of medical data in European research has been developed (Bahr et al., 2015) and deployed in the IMI-funded project eTRIKS https://www.etriks.org.

A common and reasonable concern for patients is the risk of the misappropriation of their health information, particularly of genetic data (Fallik, 2014; Millier, 1998; Riba, 2007) that may adversely affect personal circumstances, including insurance coverage and employment. (Feldman et al., 2012). Data access and confidentiality risks are directly correlated (Kum & Ahalt, 2013). In specifically excluding personal medical data from general principles making public data open-by-default, the EC's Open Data policy appears to reflect these concerns. The EC has planned to address patient confidentiality concerns through amendments to existing data protection directives (Directive 95/46/EC; EU COM/2012/11. 2012) following EU constitutional revisions that strengthened personal data protection rights (Treaty of Lisbon, 2008).

The Directive on patients' rights (Article 14) addresses legal policy cooperation on eHealth. The eHealth Network Multi-Annual Work Plan 2015-2018 defines concrete actions to be taken in the domains of interoperability and standardisation; exchange of knowledge; assessment of implementation; and global cooperation and positioning cross the European Union (Multi-Annual Work Plan 2015-2018). The adoption of guidelines on the electronic exchange of patient summaries, (November 2013) is one example of the development of electronic exchange to meet the requirements of the Cross-Border Directive (2011/21/EU. 2013: Brussels). These legislative changes unify EU initiatives on confidentiality and data security, provide a flexible legal framework that can rapidly adapt to changing technologies. Data protection reforms have arguably been few and limited to enhancing transparency and confidentiality in lawful data processing (Salas-Vega et al., 2015), but this is set to change as the GDPR comes into force in 2018 (Maldoff, 2016).

3.3 Standards & Interoperability

There are still standardization problems in the healthcare sector, as data is often stored inconsistently, or stored in systems that produce and consume incompatible formats (Roney, 2012). Research, clinical activities, hospital services, education, and administrative services are siloed, and, in many organizations. Each silo maintains its own separate organizational (and sometimes duplicated) data and information infrastructure. The lack of coordination between systems, health care providers and countries, supports the call for standards to facilitate interoperability among the components of the big data value chains (EU Report - Use of Big Data in Public Health Policy Research, 2014**)**.

3.4 Technical Issues

Robust scientific and data processing methodologies involve the validation of algorithms, filtering systems for noisy data, managing biases, and the selection of appropriate data streams. Methodological robustness is an ethical, not just a scientific requirement. Limited resources are wasted based on defective results and incomplete analyses, but also because trust in health care systems and professionals can be undermined by the use of misleading or inaccurate results (Vayena et al., 2015).

4.0 Conclusions & Future Implications

Health care delivery is a major challenge for all EU member states. The growth of big data in this sector, and the wider use of data analytics platforms offer a promise of more effective and efficient delivery of health care in the future. The MIDAS study is being undertaken to develop data forecasting, dashboards and visualisation technology to provide sustainable solutions in four European countries relating to diabetes, obesity, children in care and mental health services.
Part of the work is a systematic approach to addressing the problems of governance, confidentiality, security as well as the complex technical issues involved in working with heterogeneous data sources. The aim is to support the development of new tools for health policy development and monitoring, and ultimately new policies for health service delivery. Unless these issues can be addressed in a way which meets the legitimate concerns of health service users, and the increasingly strict legal requirements, the immense promise of Big Data for health care will remain just that, a promise.

5.0 References

Anema A, Kluberg S, Wilson K, Hogg RS, Khan K, et al. (2014). Digital surveillance for enhanced detection and response to outbreaks. *Lancet of Infectious Diseases* 14: 1035–36.

Bahr A, Schlünder I (2015). Code of practice on secondary use of medical data in European scientific research projects. *Journal of Private International Law.* 5: 279–91

Bernstein, S.W (2014). The US big data report and fully utilising big data within healthcare. *eHealth Law & Policy*, 1(4).

Brockmann D, Helbing D (2014). The hidden geometry of complex, network-driven contagion phenomena. *Science* 342: 1337–42.

Capan M, Khojandi A, Denton BT, Williams KD, Ayer T, Chhatwal J, Kurt M, Lobo JM, Roberts MS, Zaric G, Zhang S, Schwartz JS (2017). From Data to Improved Decisions: Operations Research in Healthcare Delivery, *Medical Decision Making*. 37(8):849-859.

Consolidated version of the Treaty on the Functioning of the European Union – Part 3 (2008). Union Policies and Internal Actions - Title XIV: Public Health - Article 168, *Official Journal* (115), 9. p. 122-4.

Commission, E (2012). *Proposal for a Regulation on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)* in COM 11 final: Brussels.

Commission, E (2013). *Guidelines on Minimim/Non-Exhaustive Patient Summary Dataset for Electronic Exchange in Accordance with Cross-Border Directive 21/2011EU.Brussels.*

DLA Piper (2016). Data protection laws of the world. https://www.lapiperdataprotection.com/index.html#handbook/world-map-section.

eTRIKS: European Translational Research Information and Knowledge Management Services. https://www.etriks.org

European Commission (1995). Directive on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Directive 95/46/EC.

European Commission (2012). Communication from the Commission: on the protection of individuals with regard to the processing of personal data and on the free movement of such data. COM//11.

European Commission Directorate-General for Health and Consumers Unit D3 eHealth and Health Technology Assessment. (2014). The Use of Big Data in Public Health Policy and Research. Brussels,

European Commission (2014). Communication from the Commission: towards a thriving data-driven Economy, COM/ 442.

European Commission (2014). *Multi-Annual Work Plan 2015-2018*, e.N.S.-g.o.P.f.M.-A.W.P. 2015-2018. Editor.

Fairchild AL, Bayer R (2004). Public health. Ethics and the conduct of public health surveillance. *Science* 303: 631–2.

Fallik, D. (2014). For big data, big questions remain. *Health Affairs (Project Hope)*, 33(7), 1111–1114.

Feldman B, Martin EM, Skotnes (2012). *T. Big data in healthcare:* Hype and Hope.

Georgatos F, Ballereau S, Pellet J, Ghanem M, Price N, Hood L, et al (2013). Computational infrastructures for data and knowledge management in systems biology. In: Prokop A, Csukás B, editors. *Systems Biology*. Netherlands: Springer; p. 377–97.

Hay SI, George DB, Moyes CL, Brownstein JS (2013). Big data opportunities for global infectious disease surveillance. *PLoS Medicine*.10, e1001413.

Hernandez I, Zhang Y (2017). Using predictive analytics and big data to optimize pharmaceutical outcomes. *American Journal Health System Pharmacy,* 15;74(18):1494-1500.

Hood, L., & Price, N. D. (2014). Demystifying disease, democratizing health care. *Science Translational Medicine*, 6(225), 225ed5.

Kum HC, Ahalt S (2013). Privacy-by-design: understanding data access models for secondary data. *AMIA Joint Summits Translational Science Proceedings* (1): 126-130.

Maldoff G (2016). How GDPR changes the rules for research. *The Privacy Advisor*.

Meldolesi E,van Soest J, Damiani A, Dekker A, Alitto AR, Campitelli M, et al (2016).

Standardisation data collection to build prediction models in oncology: a prototype for rectal

cancer. *Future Oncology*, London, England, (6) 12: 119-36

Manyika J, Chui M, Brown B, Buhin J, Dobbs R, Roxburgh C, Byers AH (2011). Big Data: The Next Frontier for Innovation, *Competition and Productivity*. USA: McKinsey Global Institute.

Miller, P. S. (1998). Genetic discrimination in the workplace. *The Journal of Law,* Medicine & Ethics, 26(3), 189–197.

Monteith, S., Glenn, T., Geddes, J., Whybrow, P. C., & Bauer, M. (2016). Big data for bipolar disorder. *International Journal of Bipolar Disorders*, 4(1), 10.

Neff G (2013). Why big data won't cure us. *Big Data* 1: 117–123.

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems,* 2: 3.

Rankin, D; Black, M; Wallace, J; Mulvenna, M, Bond, R; Cleland, B (2017). *The MIDAS Platform: Facilitating the Utilisation of Healthcare Big Data in Northern Ireland and Beyond*. 8th Annual Translational Medicine Conference, City Hotel, Derry/Londonderry, Northern Ireland. Clinical Translational Research and Innovation Centre (C-TRIC). 1 pp.

Riba, S. (2007). The use of genetic information in health insurance: who will be helped, who will be harmed and possible long-term effects. *Southern California Review of Law & Social Justice*, 16, 469–469.

Roney, K. (2012). If Interoperability is the Future of Healthcare, What's the Delay? Becker's Hospital Review.

Rumsfeld, J. S., Joynt, K. E., & Maddox, T. M. (2016). Big data analytics to improve cardiovascular care: promise and challenges. Nature Reviews. *Cardiology*, 13(6), 350–359.

Salas-Vega S, Haimann A, Mossialos E (2015). Big Data and Health Care: Challenges & Opportunities for Coordinated Policy Development in the EU, *Health System Reform* 1:4 285-300.

Salathé M, Khandelwal S (2011). Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLoS Computational Biology* 7: e1002199.

Thomas L (2014). Pandemics of the future: Disease surveillance in real time. *Surveillance and Society*. 12: 287–200.

Union , E.Pa.C.o.t.E., DIRECTIVE 2011/12/EU on the application of patients' rights in cross border healthcare.

Vayena E, Mastroianni A, Kahn J (2012). Ethical issues in health research with novel online sources. *American Journal of Public Health* 102: 2225–30.

Vayena E, Salathé M, Madoff LC, Brownstein JS (2015). Ethical challenges of big data in public health. *PLoS Computational Biology*, 11(2):e1003904

Vesset, D (2012). *Worldwide Big Data Technology and Services 2012–2015 Forecast.* IDC Report.

White RW, Tatonetti NP, Shah NH, Altman RB, Horvitz E. (2013). Web-scale pharmacovigilance: listening to signals from the crowd. *Journal of the American Medical Informatics Association* 20: 404–8.