Attachment G2

POLICY IMPLICATIONS OF BIG DATA IN HEALTH

Effy Vayena, University of Zurich
Joan Dzenowagis, World Health Organization
John Brownstein, Harvard Medical School
Aziz Sheikh, The University of Edinburgh

ABSTRACT

The advent of big data presents important new opportunities to improve health. New data sources, analytical models, and stakeholders increasingly interact in dynamic relationships, forming an evolving health data ecosystem, which, along with the opportunities, poses special challenges for health policy. This paper describes some of the policy challenges faced by governments, describes examples of country approaches so far, and provides directions for reshaping the policy response to facilitate the full and safe exploitation of big data for promoting health. In particular, we focus on three policy priority areas: 1) quality and safety; 2) access to data; and 3) transparency and accountability.

INTRODUCTION

Advances in health and medical knowledge depend critically on data and our ability to query, analyze, and interpret them to inform health care, research, and public health. The advent of "big data"[1] has brought important opportunities for health through the convergence of new technologies and analytical capabilities that enable rapid and broad analysis of dispersed and diverse data. While the capabilities for exploiting big data are still limited in weak health systems, in some countries the development of technical capabilities is rapidly outpacing the development of policy needed to ensure its ethical and responsible use.

THE VALUE OF BIG DATA FOR HEALTH

Health is already benefiting from the use of big data. High expectations about its potential application to clinical care, basic and translational research, personal health, and public health are fueled by enthusiasm about data analytics more generally. The big data phenomenon— frequently cited, yet vaguely defined—is credited with triggering a digital transformation of society. The "data revolution" is driven by the sheer amount of data generated daily through a myriad of channels; the speed and ease with which dispersed and diverse data can be linked, studied and shared; and the predictive power and insight that such analyses can yield. These powerful features have led to the reconceptualization of data as a new asset class, with value accruing to the data itself as well as to its uses.2 In the context of health, both the public and the private sector are investing in developing the big data ecosystem; that is, investing in the technologies and analytical capabilities that are needed to unlock the value of this important new asset. In parallel, investment in policy will be crucial to keep pace with these growing capabilities.

Linking Health Data to Create Big Data

A natural starting point for governments interested in applying big data analytics to improve

health is to link national health care data sets to facilitate in-depth analysis of health services performance and utilization, as well as the identification of other patterns. For example, in the 19 Organisation for Economic Co-operation and Development (OECD) countries using unique patient identifiers, links can be made between data sets on primary and long-term care, mental hospital in-patient care, emergency services, prescription medicine, mortality, cancer registries, and population census.[3] At the institutional level, the analyses of data from electronic health records (EHRs) greatly expand the capacity to generate new knowledge by creating an observational evidence base for clinical questions.[4,5] Big data analytics is already proving critical to building accurate models of disease progression and to realizing personalized medicine in clinical practice, such as through the integration of genomics and EHR data.[6] Other applications enable the evaluation of the impact of health policy interventions and improve the efficiency with which clinical trials are undertaken. Further, big data offers the possibility to integrate EHR data with personal data from other sources (e.g., wearable devices, sensors, medical devices, and virtual reality), thereby encouraging patients to participate in their own care, delivering personalized information directly to them and integrating medicine with behavioral determinants of health.[7]

International Efforts

Beyond the national level, international initiatives are sparking interest in cross-country collaborative projects to facilitate data sharing for health and development. For example, the Global Alliance for Genomics and Health, a network of over 400 institutions in health care, research, advocacy, life science, and information technology, is developing policies and technological solutions to facilitate genomic data sharing.[8] Similarly, the European Medical Information Framework seeks to develop a model to connect health data from a variety of sources to facilitate large-scale research.[9] Notably, development partners have explored how to use big data in sustainable development, including in international public health, but many of the countries that could benefit most are in need of considerable infrastructure and capacity development. The adoption of the United Nations Sustainable Development Goals emphasized the role of measurement and accountability in development, and both donor and recipient governments are poised to accelerate investment in big data capabilities to support this goal.[10]

**Big Data Beyond the Health Sector**

A vast amount of data is generated outside the health care system, yet can be used for health research, particularly with respect to studying personal and group characteristics, behaviors, and interactions. The data-rich online environment includes the Internet, web, and their platforms (e.g., social media), services (e.g., purchasing, email), and applications.[10] Such data are considered an important asset by the commercial sector for consumer profiling and marketing

purposes, as incorporating the digital "exhaust" or traces created by consumer activity can provide insights that would not otherwise be possible. Health research with these nontraditional data streams has already demonstrated its value.[12,13] One recent illustration is the Zika virus, which emerged in the Americas. The event exposed the lack of basic epidemiological information on the disease. However, the availability of online media reports made it possible to supplement existing information and close the knowledge gap, allowing researchers to estimate Zika virus transmission dynamics and obtain the insight needed to plan response measures beyond vector suppression.[14]

The Big Data Ecosystem

The above examples illustrate the continuous expansion of the data types, sources, and possibilities of their use. The new opportunities for analyzing data within and across sectors for health purposes have various effects. For example, they allow an alternative approach to the standard biomedical research paradigm that starts with a research hypothesis and aims to collect evidence to refute or confirm it. They also encourage a more comprehensive view of health, whereby all data that are captured about, or from, a person can contribute to understanding biology, health, and disease using a data mining approach. Given this scope, the stakeholders extend beyond the patient/consumer, health care providers, and research and academic institutions to include businesses, professional societies, government and development agencies, and other entities not necessarily directly related to health research or the delivery of health services. Therefore, as new data sources, analytical models, and stakeholders increasingly interact within dynamic relationships, it is helpful to think of big data in health as an ecosystem (Figure 1).

Figure 1. The evolving health data ecosystem shows the expanding health data sources [green], the increasing capabilities enabling data investigation and use [violet], and the diversity of stakeholders [blue] that are creating new opportunities for health.
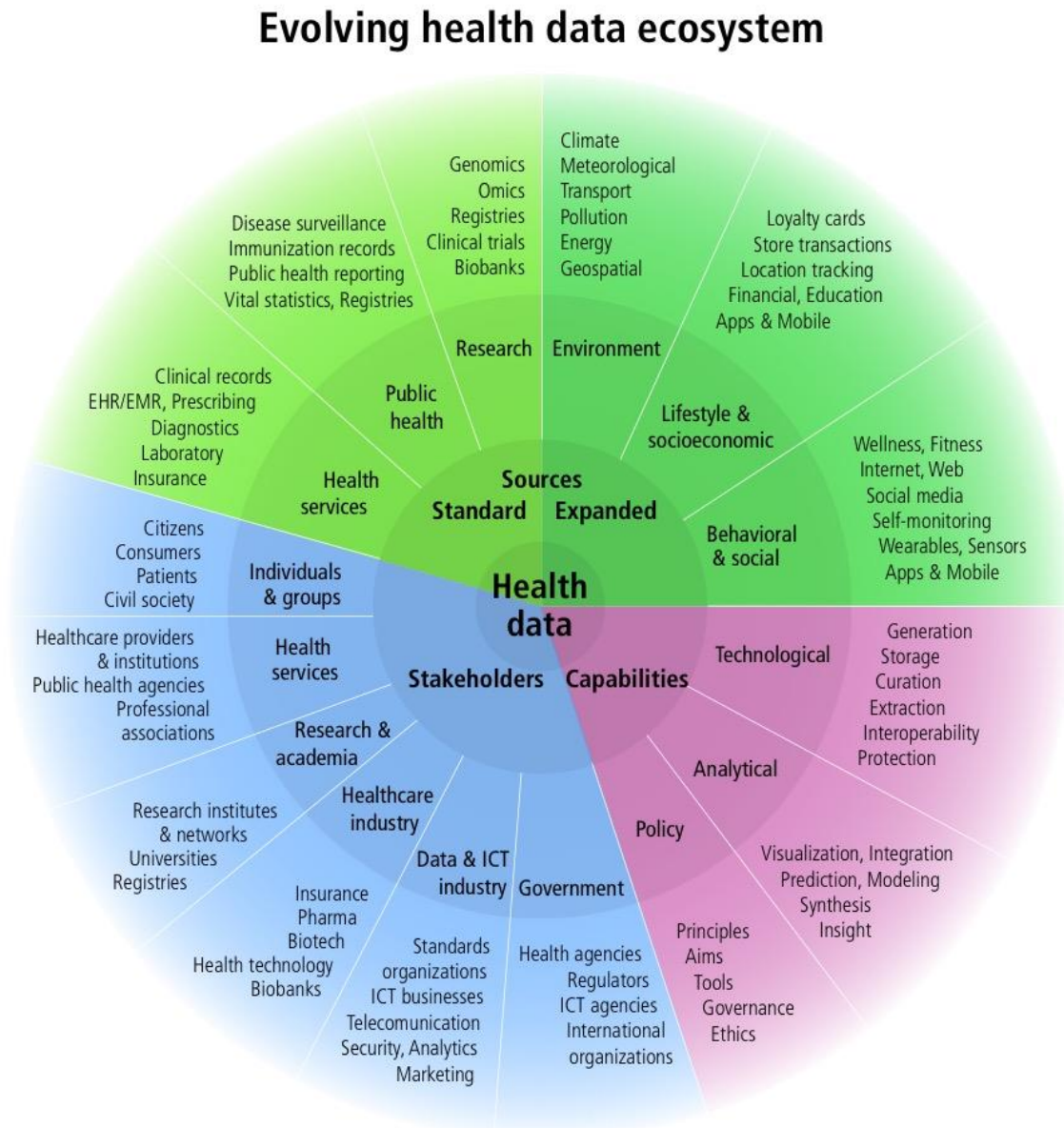
E. Vayena, J. Dzenowagis, M. Langfeld, 2016

POLICY CHALLENGES IN BIG DATA USES FOR HEALTH

This ecosystem, with increasing data sources, new capabilities, and an expanded set of stakeholders poses special challenges, which continue to evolve. Governments need to consider

how to reshape national policies, not only to advance and improve the use of big data in health, but also to take into account the characteristics of big data and its implications for the public interest. In particular, the privacy, security, and confidentiality of health-related data remain an ongoing concern. Big data methods introduce new risks, due to what can be inferred from the data as a whole or when linked to other information. Risks may arise from the characteristics, scope, and combination of the data; the systems, policies, and technologies to manage it; and the broader context of its use. Even basic health data of vulnerable populations can be misused and lead to discrimination. Conversely, a range of new benefits can arise from data exploitation, which poses further challenges as to how these benefits can be fairly distributed. The following sections present some specific challenges.

Data Quality

Data reliability and quality are commonly cited challenges to effective data utilization.[15] Reliability can be compromised by data entry errors (for example, in manually fed systems) or by systemic bias introduced in the data set—for example, if data are from a population subgroup only or due to algorithmic flaws in the processing. Data identification and methods used for data science and open data sets may have degraded quality because they are derived, rather than primary, data sources. This can potentially affect the research results obtained from the inaccurate interpretation of third-party data sources. Other data science methods that have been foundational to assuring quality may no longer be useful due to the combination of data sets that were never intended to be analyzed together. For example, assumptions about the data may not hold true across different data sets, leading to invalid interpretation.

Interoperability

Integrating data from diverse sources in the health sector is greatly aided by interoperable systems, which must integrate with information architectures that may not have been designed for a big data scenario. Developing an interoperable health information architecture to handle dissimilar data sets and heterogeneous data is a major challenge. It requires agreement on the data items to be shared, a common language and standards, and a secure means of sharing. Incorporating non-health care data increases the complexity still further.

Data Access and Sharing

Secondary uses of data held by an institution often remain the exclusive privilege of the institution's research team. Data sharing with other organizations is generally limited. OECD devised a set of criteria that determine the level of health data accessibility in countries. These include whether identifiable data are shared with other institutional data custodians and government entities, and whether de-identified data can be approved for access by universities

and nonprofit research entities or by for-profit businesses and foreign government organizations. A survey of countries revealed that only a few meet most accessibility criteria, while many lack any mechanism for data access and use.3

Unwillingness to share data is often attributed to concerns about privacy, consent, or security.[16] However, even when these concerns can be addressed, organizations or institutions that control data may have little incentive to share them. Storing, maintaining, and securing data sets is costly, and therefore institutions are naturally inclined to exploit the data for their own benefit— be it publication of research results, monetization, or other purposes. It may also be difficult to come to agreement as to partner accountability, responsibilities, and rights to intellectual property resulting from data sharing. There is currently no comprehensive framework within which benefits to stakeholders are defined and global health data sharing can take place. In the global public health setting, these limitations came sharply into focus during the 2014-2015 Ebola emergency. Despite the massive amounts of data collected by the health sector, data were not available for further research.[17,18] The international community reacted with a call for action that resulted in the creation of a protocol for immediate online posting of data where researchers could share their data while retaining authorship and precedence.[19] More recently, an international group led by the Chatham House called for the need for data sharing for routine global health surveillance, not only during emergency situation.[20]

Privacy

Health data are sensitive because they are personally identifiable and must be afforded confidentiality and privacy protections. Privacy is understood here as an individual's control of their personal information. A typical measure of privacy in the health sector is that data are obtained—with consent—for a particular purpose, and secondary uses beyond that specific purpose may be limited or impossible. In recent years, different consent models have emerged aiming to enhance data control and utilization.[21, 22] However, as new uses and applications are increasing, consent is no longer sufficient as a privacy control.[23] Consent cannot cover all possible uses of data, which exponentially increase as data sets become linked. For example, concerns about medication safety may emerge long after cohorts have been established, making it impossible to obtain individual consent when undertaking pharmacovigilance review.[24] Purpose-specific consent, which can be lost in the complexity of the big data environment, will need to transition to a more transparent and effective means of data governance. Anonymization is heavily relied on as a privacy protection measure. In many instances, an individual's anonymized data can be used for secondary purposes without consent but with approval by an institutional review board. However, in the big data environment, anonymization presents serious vulnerabilities. Re-identification remains a risk due to the technology features of data analytics

and the possibilities of data combinations.[25,26] Researchers have identified individuals in a publicly available genomic database stripped of identifiers by combining the data with recreational genealogy data and other data such as age and state.[27] In light of the above, the benefits and risks of data sharing will need careful consideration in the next generation of risk minimization of data sharing and appropriate use.

Data Ownership

Although the concept of "data ownership" is highly debated, in practice data controllers often act as data owners.[28] This can be at odds with the idea that the actual data owners may be the individuals from whom the data have been generated. There is wide variation in how regulators in different jurisdictions approach the issue of health data ownership. In some countries, health data from within the health care system are by law the property of patients (e.g., Switzerland), while in others the data are the property of the health care system. This diversity adds further complexity in international data sharing projects. Increasingly, the private sector possesses large volumes of health data. Access to health data controlled by the private sector is even more challenging because, for many such businesses, data are their major asset. Some business models involve the monetization of health data (typically in anonymized form). For example, companies such as 23andMe, a direct-to-consumer genomic testing company, has amassed data from more than one million people.[29] The company sells access to its data pool to pharmaceutical companies. The price tags of data access in such business models can be a limiting factor to broader access, e.g., by publicly funded research organizations. Private companies may also control non-health data, which, with new analytic capabilities, can be used for health purposes. This control raises questions regarding the responsibility of such companies, for example, in public health emergencies. This issue came to the fore during the 2014-2015 Ebola virus crisis in West Africa, when privately held telecommunication data were needed to assist public health agencies to track the epidemic.[30]

COUNTRY APPROACHES TO BIG DATA IN HEALTH

The use of big data in the health sector depends in large part on the health systems development context. While there is increasing awareness of the benefits of big data, and technological and methodological progress is being made, there is still a lag in the adoption of big data methods in countries. Reasons include gaps in leadership, technical expertise, and funding, as well as competing priorities for the health system.[31] Another critical reason is that governments are still weighing appropriate policy options.

According to the World Health Organization Global Observatory for eHealth, in 2015 less than a fifth of countries (n=21, 17%) reported having a policy or strategy regulating the use of big data

in the health sector.[32] When analyzed by World Bank income group, nearly a quarter of countries (n=7, 23%) in the upper-middle income group had a national policy or strategy regulating the use of big data in the health sector. Differences among the other three World Bank income groups were minimal: high income (16%), lower-middle income (14%), and low income (16%) countries respectively.[32]

Figure 2. Percent of Countries with a National Policy or Strategy Regulating the Use of Big Data in the Health Sector (n=121)[33]
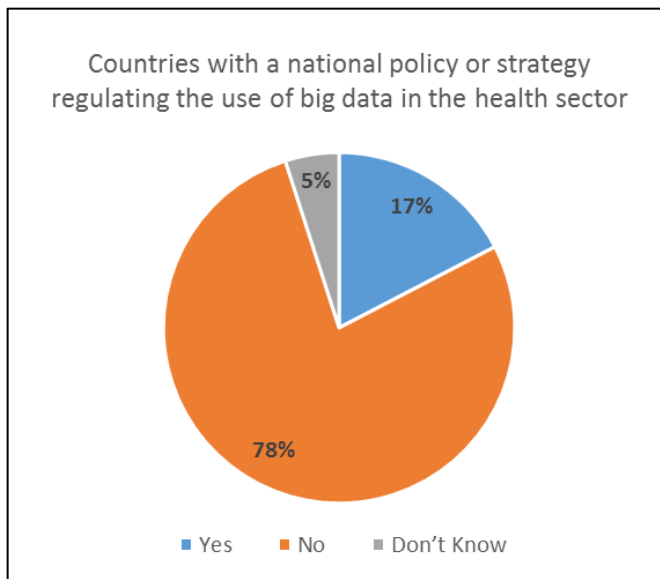
Source: World Health Organization, 2016.

A number of countries have been working toward the systematic utilization of big data in health. A major prerequisite for data utilization is a certain level of information and communication technology infrastructure and standardization. Also needed are sound and up-to-date capabilities in data science and policies that facilitate and support the use of big data in the public interest. The examples below illustrate some of the approaches taken.

Scotland

Scotland's Community Health Index is a national population register through which people are uniquely identified using a 10-digit number. This standard allows health care data sets to be linked and increasingly allows health care and non-health care data to be linked on the overwhelming majority of the Scottish population. These linked data sets are used for clinical purposes, as well as to support research using Scotland's extensive national data sets. Scotland, along with England and New Zealand, are among the OECD countries with highest data

accessibility scores. Considerable care is taken to ensure the security using a principled proportionate governance approach, based on the risks identified. Data use is encouraged for patient and population benefit where data are already in the public domain. Where the risks of disclosure and consequences are deemed high, then a range of safeguards are put in place, including full ethical and privacy review, ensuring that researchers are appropriately trained and undertaking data analysis in trustworthy research environments. This approach has the benefits of balancing both the imperative to promote use of data to support health care delivery and research and minimization of the risk of undermining an individual's privacy. Examples of using this approach include a national evaluation investigating the effectiveness of the H1N1 vaccination in the context of preventing pandemic influenza, which involved linking primary care, hospital, virological, and mortality data.[34] A further example was the assessment of ethnic and racial variations in hospitalization, readmission, and mortality from asthma on over 90 percent of the Scottish population using linked census, hospital, and mortality data.[35] Data linkage was necessary because ethnicity data were poorly recorded in health data but were nearly completely recorded in census data. Similarly, linking disparate data sets has made it possible to assess the impact of the ban on smoking in enclosed public places on asthma hospital admissions.[36]

Denmark

Denmark has been a pioneer in e-health and is considered an example for countries to follow. A key agenda of the health system is to ensure the availability of accurate and easily accessible personal health information, while protecting privacy. Important features of the Danish system include a unique electronic personal identifier from birth and a shared medical card with encoded prescriptions information to which relevant health care practitioners have access. The sundhed.dk portal makes citizens' own health data accessible to them and their health care providers. Denmark's 95 percent broadband penetration (the highest in Europe) and high literacy rate make it easy for citizens to access their data and the e-health system, including from mobile devices. Data linkage is possible at various levels, and Danish personal health data sets meet most of the OECD accessibility criteria. One recent example of big data utilization with important policy implications is a study of hospital bed occupancy that revealed a strong association between hospital bed shortage and increased mortality.[37] Denmark's infrastructure and policy facilitates big data use within a robust privacy environment, with a high level of public scrutiny. For example, a national clinical database for primary care data established to promote quality improvement in general practice and enable research on primary care was deleted in 2014. In response to public concerns, the National Board for e-health determined that the database would pose privacy risks that threatened to damage trust between general practitioners and their patients.[38,39] This case highlights the significance of public engagement on

the issue of privacy and the priority of maintaining the public's trust.

England

In 2013, England's National Health Service launched the care.data initiative to place patient data from general practitioners around the country in the Health and Social Care Information Center database (HSCIC; now NHS Digital). HSCIS already collected hospital data. Analyzing data from general practitioners would not only improve outcomes and customer service, but would also improve understanding of diseases and treatments. Although care.data was initially endorsed by various professional societies, it eventually stalled after negative public response. Concerns about privacy, lack of transparency regarding data access, and the involvement of commercial entities led to a strong public reaction against the initiative. The ensuing national and international debate spotlighted the thorny issues of big data use and more specifically the issue of trust. Care.data was finally discontinued following reports by the National Data Guardian and the Care Quality Commission that highlighted inadequacies in transparency and privacy protection.[40] The report emphasized that to build public trust in the use of health and care data, citizens should be able to exercise their "right to know how their data is safeguarded. They should be included in conversations about the potential benefits that responsible use of their information can bring. They must be offered a clear choice about whether they want to allow their information to be part of this." Undoubtedly, big data initiatives need to meet criteria of trustworthiness, transparency, open communication, and a clear sense of the distribution of benefits to gain the trust of the public.[41]

POLICY DIRECTIONS

Governments adopting big data technologies and methods in the health sector will need to: establish proactive and durable policies in the public interest; protect the health data of individuals (privacy, security, confidentiality); tackle the reality of its commercialization; and further the interoperability and use of public data created and held as a public good. In particular, the focus should be on three priority areas: 1) quality and safety; 2) access to data; and 3) transparency and accountability.

Quality and Safety

The information systems within which health data are generated, stored, managed, and used need to meet high quality standards. Such standards are evolving, along with technological developments and emerging quality challenges. Policies can mandate or provide incentives for standardization and system optimization to meet quality and safety goals.[42] At the national level, data integration strategies, interoperability, and security standards should be agreed upon. This should include certification and compliance for the information and communication technologies

infrastructure for safety and security. Setting standards can engage industry stakeholders, boost competition and quality, and provide more options for big data clients to choose from. The latter can facilitate adoption of EHRs as seen in the United States, Denmark, and elsewhere.

Incidents of data breaches and "kidnapping" (data held by hackers for ransom) are on the rise. According to the Breach Portal of the U.S. Department of Health and Human Services (HHS) Office for Civil Rights, millions of health care records have been affected to date. Notable cases in 2016 include the Hollywood Presbyterian Medical Center for which a bitcoin ransom was requested and eventually paid in return for security keys that unlocked the affected devices. A similar case involved the Kansas Heart Hospital, and a recent breach was reported by Massachusetts General Hospital.[43] The United Kingdom's Information Commissioner's Office notes that the health sector accounts for most of the data incidents reported to them.**Error! Reference source not found.** Against this background, the public needs to be reassured that robust security measures are mandated and enforced through clearly articulated policies. Concerns can be addressed with the adoption of appropriate technologies, monitoring and evaluation of security systems, and transparency and accountability mechanisms such as legal remedies and compensation for those harmed by security breaches. Security as a technological and societal norm will continue to evolve, and the big data approach will continue to demand more technical skills, responsive policies, and regulatory oversight.

Data Access and Benefit Sharing

Health data exploitation is predicated upon data being accessible to those who will use them in the public interest. This requires a comprehensive data governance framework that spells out: conditions of data access, including appropriate safeguards; roles and responsibilities of data users; and principles of benefit sharing.[26] Past frameworks placed most ethical control of data use at the point of data generation such as through consent. However, the big data environment with its infinite possibilities of data combinations and uses requires controls throughout the life cycle of data and at the point of use. A key element here is clarity on privacy protections that apply at different types of use and across the health data ecosystem.

Privacy is not an "on/off" concept, and its many gradients require the development of corresponding protections. Proportionate and effective privacy safeguards can facilitate data access. New privacy technologies will play a crucial role in a robust privacy approach. Legal reforms on privacy and data protection are underway in many countries, e.g., the European Union General Data Protection Regulation and revision of the Swiss Federal Data Protection Act. Such reforms attempt to increase data subjects' privacy choices and introduce controls on data uses. While legal reforms intend to bring clarity, they may also introduce rigidity that is

unlikely to serve the evolving and multifaceted big data ecosystem.[45] Ideally, reforms will be forward looking with enough flexibility to allow for quick adaptation. However, it should be noted that creating rules for data access is not just a matter of data protection policy. The distribution of benefits of data exploitation is a determining factor of data access, along with the public acceptability of the use of personal data.[46] Policies should therefore seek to establish the principles for fair benefit sharing as an integral part of data access.

Moreover, it is important that policies aiming to support health data access are citizen-centric. Typically, data access refers to access by third parties, often failing to address access by citizens themselves. Citizens increasingly demand access to their own data partly to control secondary uses.[47,48] Technologies can facilitate individual data access, and individuals themselves may provide the "missing link" in data access for the entire health data ecosystem.[49] Citizens are stakeholders in the development, implementation, monitoring, and evaluation of data initiatives. Therefore, their role should extend far beyond providing informed consent for data uses. They should be sufficiently consulted and have a meaningful role in the governance of data initiatives. This augmented role can also facilitate the negotiation of fair benefit sharing.

Transparency and Accountability

As the health data ecosystem continues to expand in tandem with increasingly sophisticated and complex methods of data linkage and use, data transactions risk becoming even more opaque. This major concern in the wider big data phenomenon is especially relevant in the health sector, where transparency and accountability are pillars of trust. Policies need to focus on maintaining a high level of transparency, particularly on the uses of big data (e.g., through data traceability) and the benefits resulting from use. Developing suitable accountability and transparency approaches should be part of an iterative process involving all stakeholders. Traditional actors in the health sector have over time developed accountability processes that need to be renewed and extended to include new actors (e.g., data and social media companies) and new capabilities (e.g., black box medicine) and cover anyone involved in the handling of personal data. This is particularly important when considering the role that new actors and specifically non-state actors play in the creation of big data norms. This quasi-legislative role arises partly from the fact that powerful non-state actors are often in possession of the data sets, have special control and expertise regarding their handling (e.g., encryption), and operate globally and hence are able to influence standard-setting across a variety of jurisdictions.

With the ethics controls of big data focusing on data uses and their purpose, sound accountability mechanisms help monitor compliance with ethical norms such as privacy protection and fairness in benefit sharing and provide avenues for seeking appropriate remedies in the case of failure. As

machine learning (computers acting without being explicitly programmed) finds its way into health care, it will be difficult to determine how decisions have been made by computers. In anticipation of these developments, policies should demand that accountability is built into the design of such systems and that algorithmic decisions are critically questioned and closely monitored. Both the recent White House report *Big Data: A Report on Algorithmic Systems, Opportunity and Civil Rights* (2016) and the European Data Protection Supervisor's Opinion *Meeting the Challenges of Big Data* (2015) rightly emphasized the central role of accountability in harnessing the potential of big data.[50,51]

CONCLUSIONS

Big data can yield significant benefits for individual and population health. However, the effective exploitation of big data toward this goal is entangled with technological, ethical, legal, and societal challenges. Policies must be designed to take into account the distinct features of these challenges in light of the magnitude of the potential benefit. They also need to be applicable to the full range of stakeholders, not least the individual citizens, and must be accompanied by a level of accountability that, over time, is sufficient to maintain the public's trust and confidence in how data are used. In this article we sketched three priority areas for policy on the basis of which a comprehensive policy framework for big data in health can be built.

References

1. Overview paper.

2. World Economic Forum, *Personal Data: The Emergence of a New Asset Class* (Geneva: World Economic Forum, 2011), http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf. Accessed July 18, 2016.

3. Organisation for Economic Co-operation and Development, *Health Data Governance: Privacy, Monitoring and Research – Policy Brief* (Paris, France: OECD Publishing, Oct. 2015). https://www.oecd.org/health/health-systems/Health-Data-Governance-Policy-Brief.pdfw. Accessed July 18, 2016.

4. Ayers J, Althouse B, Dredze M, "Could Behavioral Medicine Lead the Web Data Revolution?" *JAMA*, 2014 311(14):1399.

5. Ayers J, Westmaas J, Leas E, et al., "Leveraging Big Data to Improve Health Awareness Campaigns: A Novel Evaluation of the Great American Smokeout," *JMIR Public Health and Surveillance*, 2016 2(1):e16. http://publichealth.jmir.org/2016/1/e16/.

6. Jensen PB, Jensen LJ, Brunak S, "Mining Electronic Health Records: Towards Better Research Applications and Clinical Care," *Nature Reviews Genetics*, 2012 13(6):395–405.

7. Kish LJ, Topol EJ. "Unpatients – Why Patients Should Own Their Medical Data," *Nature Biotechnology*, 2015 33(9):921–24.

8. Global Alliance for Genomics & Health, Toronto, http://genomicsandhealth.org. Accessed July 18, 2016.

9. European Medical Information Network, http://www.emif.eu/. Accessed July 18, 2016.

10. Sustainable Development Solutions Network Leadership Council, *Indicators and a Monitoring Framework for the Sustainable Development Goals: Launching a Data Revolution for the SDGs* (May 15, 2015), http://unsdsn.org/wp-content/uploads/2015/05/FINAL-SDSN-Indicator-Report-WEB.pdf. Accessed July 18, 2016.

11. Jain S, Powers B, Hawkins J, Brownstein J, "The Digital Phenotype," *Nature Biotechnology*, 2015 33(5):462–63.

12. McIver D, Brownstein J, "Wikipedia Usage Estimates Prevalence of Influenza-Like Illness in the United States in Near Real-Time," *PLoS Computational Biology*, 2014 10(4):e1003581. http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003581.

13. Brownstein J, Freifeld C, Madoff L, "Digital Disease Detection – Harnessing the Web for Public Health Surveillance, New England Journal of Medicine," 2009 360(21):2153–57.

14. Majumder MS, Santillana M, Mekaru SR, McGinnis DP, Khan K, Brownstein JS, "Utilizing Nontraditional Data Sources for Near Real-Time Estimation of Transmission Dynamics During the 2015-2016 Colombian Zika Virus Disease Outbreak," *JMIR Public Health and Surveillance*, 2016 2(1):e30. http://www.ncbi.nlm.nih.gov/pubmed/27251981.

15. Mossialos E, Wenzl M, Osborn R, Sarnak D, eds, *2015 International Profiles of Health Care System*s (New York, N.Y.: The Commonwealth Fund, Jan. 2016). http://www.commonwealthfund.org/~/media/files/publications/fund-report/2016/jan/1857_mossialos_intl_profiles_2015_v7.pdf. Accessed July 18, 2016.

16. van Panhuis WG, Paul P, Emerson C, et al., "A Systematic Review of Barriers to Data Sharing in Public Health," *BMC Public Health*, 2014 14:1144.

17. Yozwiak N, Schaffner F, Sabeti PC, "Data Sharing: Make Outbreak Research Open Access," *Nature*, 2015 518(7540):477–79.

18. Delaunay S, Kahn P, Tatay M, Liu J, "Knowledge Sharing During Public Health Emergencies: From Global Call to Effective Implementation," *Bulletin of the World Health Organization*, 2016 94(4):236–36A.

19. Dye C, Bartolomeos K, Moorthy V, Kieny MP, "Data Sharing in Public Health Emergencies: A Call to Researchers," *Bulletin of the World Health Organization*, 2016 94(3):158.

20. Chatham House, The Royal Institute of International Affairs, Strengthening data sharing for public health, https://www.chathamhouse.org/about/structure/global-health-security/strengthening-data-sharing-public-health-project. Accessed July 18, 2016.

21. Vayena E, Mastroianni A, Kahn J, "Caught in the Web: Informed Consent for Online Health Research," *Science Translational Medicine*, 2013 5(173):173fs6.

22. Kaye J, Whitley EA, Lund D, Morrison M, Teare H, Melham K, "Dynamic Consent: A Patient Interface for Twenty-First Century Research Networks," *European Journal of Human Genetics*, 2015 23(2):141–46.

23. Vayena E, Gasser U, "Between Openness and Privacy in Genomics," *PLoS Medicine*, 2016 13(1):e1001937. http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001937.

24. Kotz D, Viechtbauer W, Simpson C, van Schayck OC, West R, Sheikh A, "Cardiovascular and Neuropsychiatric Risks of Varenicline: A Retrospective Cohort Study," *The Lancet Respiratory Medicine*, 2015 3(10):761–68.

25. de Montjoye YA, Radaelli L, Singh VK, Pentland AS, "Identity and Privacy. Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata," *Science*, 2015 30 (6221):536–39.

26. Vayena E, Gasser U, Wood A, O'Brien D, Altman M, "Elements of a New Ethical Framework for Big Data Research," *Washington and Lee Law Review*, 2016 72(3):420–41.

27. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y, "Identifying Personal Genomes by Surname Inference," *Science*, 2013 339(6117):321–24.

28. Evans BJ, "Much Ado About Data Ownership," *Harvard Journal of Law & Technology*, 2011 25(1):70–130.

29. 23andMeBlog, Power of one million, https://blog.23andme.com/news/one-in-a-million/. Published June 18, 2015. Accessed July 18, 2016.

30. Wesolowski A, Buckee CO, Bengtsson L, et al., "Commentary: Containing the Ebola Outbreak – The Potential and Challenge of Mobile Network Data," *PLoS Current Outbreaks*, 2014 29(6). http://currents.plos.org/outbreaks/article/containing-the-ebola-outbreak-the-potential-and-challenge-of-mobile-network-data/. Published Sept. 29, 2014. Accessed July 18, 2016.

31. World Health Organization and International Telecommunication Union, National eHealth strategy toolkit, Geneva, https://www.itu.int/dms_pub/itu-d/opb/str/D-STR-E_HEALTH.05-2012-PDF-E.pdf. Published 2012. Accessed July 18, 2016.

32. World Health Organization, Global observatory for eHealth, Geneva, http://www.who.int/goe/en/. Published 2015. Accessed July 18, 2016.

33. World Health Organization, The third global survey on eHealth, Geneva, 2016, to be available at www.who.int/goe/publications.

34. Simpson CR, Ritchie LD, Robertson C, Sheikh A, McMenamin J, "Effectiveness of H1N1 Vaccine for the Prevention of Pandemic Influenza in Scotland, UK: A Retrospective Observational Cohort Study," *The Lancet Infectious Diseases*, 2012 12(9):696–702.

35. Sheikh A, Steiner MFC, Cezard G, et al., SHELS researchers, "Ethnic Variations in Asthma Hospital Admission, Readmission and Death: A Retrospective, National Cohort Study of 4.62 Million People in Scotland," *BMC Medicine*, 2016 14(3).

36. Mackay D, Haw S, Ayres JG, Fischbacher C, Pell JP, "Smoke-Free Legislation and Hospitalizations for Childhood Asthma," *New England Journal of Medicine*, 2010 363(12):1139–45.

37. Madsen F, Ladelund S, Linneberg A, "High Levels of Bed Occupancy Associated with Increased Inpatient and Thirty-Day Hospital Mortality in Denmark," *Health Affairs (Millwood)*, 2014 33(7):1236–44.

38. Mossialos E, Wenzl M, Osborn R, Sarnak D, eds, *2015 International Profiles of Health Care Systems*, (New York, N.Y.: The Commonwealth Fund, Jan. 2016).

39. Christiansen T, Rudkjøbing A, Deletion of the Danish General Practice Database (DAMD), http://hspm.org/countries/denmark27012013/livinghit.aspx?Section=2.7%20Health%20information%20management&Type=Section#12DeletionoftheDanishGeneralPracticeDatabase(DAMD). Accessed July 18, 2016.

40. Freeman MP, Gov.UK, Department of Health, Review of health and care data security and consent, https://www.gov.uk/government/speeches/review-of-health-and-care-data-security-and-consent. Published July 6, 2016. Accessed July 20, 2016.

41. Carter P, Laurie GT, Dixon-Woods M, "The Social Licence for Research: Why Care.data Ran into Trouble," *Journal of Medical Ethics*, 2015 41(5):404–09.

42. Sheikh A, Jha A, Cresswell K, Greaves F, Bates DW, "Adoption of Electronic Health Records in UK Hospitals: Lessons from the USA," *The Lancet*, 2014 384(9937):8–9.

43. Department of Health and Human Services, Office of Civil Rights, Breaches affecting 500 and more individuals, https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf. Accessed July 20, 2016.

44. Information Commissioner's Office, United Kingdom, Data security trends, https://ico.org.uk/action-weve-taken/data-security-incident-trends/. Accessed July 20, 2016.

45. Sethi N, Laurie GT, "Delivering Proportionate Governance in the Era of eHealth," *Medical Law International*, 2013 13(2-3):168–204.

46. Ipsos MORI Social Research Institute for the Wellcome Trust, *The One-Way Mirror: Public Attitudes to Commercial Access to Health Data*, https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf. Published March 2016. Accessed July 18, 2016.

47. Pentland A, Reid TG, Heibeck T, for the Big Data and Health Working Group 2013, *Big Data and Health: Revolutionizing Medicine and Public Health*, World Innovation Summit for Health, http://www.wish-qatar.org/app/media/382. Published December 2013. Accessed July 18, 2016.

48. Hafen E, Kossmann D, Brand A, "Health Data Cooperatives—Citizen Empowerment," *Methods of Information in Medicine*, 2014 53:82–86.

49. Weber GM, Mandl KD, Kohane IS, "Finding the Missing Link for Big Biomedical Data," *JAMA*, 2014 311(24):2479–80.

50. Executive Office of the President, *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights* (Washington, D.C.: May 2016), https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf. Accessed July 18, 2016.

51. European Data Protection Supervisor, Opinion 7/2015, *Meeting the Challenges of Big Data* (Brussels, Belgium: Nov. 19, 2015), https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2015/15-11-19_Big_Data_EN.pdf. Accessed July 18, 2016.