Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

Ethics and Big Data in health

Bartha Maria Knoppers and Adrian Mark Thorogood

Abstract

With the EU General Data Protection Regulation entering into force in 2018, the stage is set for international debate on Big Data sharing in health. Considering the fact that health data (and especially genetic data) are considered "sensitive", is there a way to structure the debate on the barriers, and risk-benefit ratio that moves away from the traditional pros and cons of potential privacy and discrimination risks? Potential discrimination has been addressed in legislation and the balancing of privacy rights against the potential benefits of data sharing in intensive science is leading to a more proportionate approach. We postulate that an important catalyst that will further shift this debate from its traditional contours would be to activate the "right to science" as enshrined in international law. Of note, the Global Alliance for Genomics and Health has developed a *Framework for Responsible Sharing of Genomics and Health-Related Data* based on that human right. Similarly, recent guidelines from the Council for International Organizations of Medical Sciences and the Organisation for Economic Co-operation and Development, as well as from the American College of Medical Genetics and Genomics contain provisions that promote data sharing indicating that data intensive science may gradually come to be founded on a more communal ethos.

Addresses

Centre of Genomics and Policy, McGill University, Montreal, Québec, H3A 0G1 Canada

Corresponding author: Knoppers, Bartha Maria (bartha.knoppers@mcgill.ca)

Current Opinion in Systems Biology 2017, **4**:53–57

This review comes from a themed issue on **Big data acquisition and analysis (2017)**

Edited by **Pascal Falter-Braun and Michael A. Calderwood**

For a complete overview see the Issue and the Editorial

Available online 8 July 2017

http://dx.doi.org/10.1016/j.coisb.2017.07.001

Keywords

Big Data, Data sharing, Genomics, Ethics, Law, Privacy, Consent.

Introduction

For clarity, best to begin with definitions. "Ethics is about goods that we have reason — and sometimes even an obligation — to pursue, such as the good of knowledge that can be used to bring about significant improvements in health" [1]. Big Data in health can be defined as "encompassing high volume, high diversity biological, clinical, environmental, and lifestyle information collected from single individuals to large cohorts, in relation to their health and wellness status, at one or several time points" [2].

Within the multidisciplinary contexts of data sharing that reinforce and enable both discovery and infrastructure science [3], the recent scale of data collection and use in biomedical research and clinical care is seemingly limitless in the quest for precision medicine via the use of next generation sequencing technologies. Yet, as genomic data enters the clinic will the data sharing ethos that characterized the Human Genome Project become part of the clinical culture as well? [4] In other words, will health data flow to and from medical record to the research context and back in a learning healthcare system? [5] With the entry onto force in 2018 of the European Union's 2018 General Data Protection Regulation (GDPR), the stage is set for an international debate on the use of Big Data in both research and health care [6].

It is not just the ethical imperative to pursue "the good" or the volume of Big Data, but its multivariate nature that inspires us to re-examine the "classical" socio-ethical issues surrounding the risks and benefits of data collection, access, and sharing and their impact on privacy and discrimination (i). Perhaps it is time to move to a new paradigm where we catalyze and activate the right of all citizens to benefit from advances in science via data sharing, its benefits, and its applications as probabilistic at-risk health information becomes the "treatment" (ii). If so, while consent and privacy remain central, governance policies and security mechanisms for Big Data will emerge as equally important (iii).

(i) Balancing Risks and Benefits

The two decades since the completion of the Human Genome Project have seen a variety of legislative and policy responses to the presumed and real socio-economic and privacy risks spawned by an increase in genetic data research, biobanks, and ensuing databases [7]. As concerns the privacy of such data and possible genetic discrimination, it bears noting that even a country with universal health care coverage such as Canada has in 2017 adopted legislation prohibiting the use of genetic data by life insurers, following the lead of European countries [8]. Consolidating decades of concern, the Council of Europe recommended in 2016 that for insurance purposes genetic tests neither be required nor used [9]. The United States adopted the

Genetic Information Nondiscrimination Act (GINA) as concerns health and employment insurance in 2008.

Another area of concern has centered on consent to the use of health data including genetic data. The emergence of bioresources in the form of large, national population cohorts creating databases for future unspecified research with ethics approval and governance led to a decade of academic debate on the legality of such broad consent, even though it typifies the longitudinal and epidemiological nature of biobanking. Today, broad consent is recognized in the Council for International Organizations of Medical Sciences (CIOMS) *International Ethical Guidelines on Health-Related Research* [10] and in the United States Common Rule [11]. Fear of misuse has to some extent been mitigated by genetic discrimination legislation and the development of sophisticated security mechanisms, but privacy concerns remain. Database may be subject to law enforcement access and mass government surveillance. Moreover, whole genome sequencing and Big Data are likely to reveal unanticipated information as interpretation improves, and in turn unanticipated third party misuses [12]. Irrespective, regulatory frameworks may not be able to track or respond to all instances of discrimination, or may not apply to new forms of predictive health information. The conditions of sharing with for-profit companies, though necessary for translation of knowledge into health interventions, remain controversial [13]. Yet, there is also the argument to be made that Big Data will be useful to sustain and improve health care and health care systems as it permits both stratification of services and testing and targeted resource allocation.

(ii) The "right to science" in international law

We maintain that, much as there is vulnerability and a reasonable apprehension of misuse of Big Data, it is on the sanctioning of the latter and not on the stifling of the beneficial avenues of the use of Big Data that we should deploy our efforts [6]. To do so successfully, it is time to activate a human right that hitherto has lain largely dormant – the right of everyone "to share in scientific advancement and its benefits". This human "right to science" has its origins in the 1948, Universal Declaration of Human Rights [14], and was made legally binding under the International Covenant on Economic, Social and Cultural Rights of 1966, signed and ratified by 165 countries [15].

Because of its public international law status, the content of this human right has universal force and its "actionability" can reach beyond the moral appeals of bioethics. It imposes positive duties on States [16]. Until now there have been limited efforts to develop the content of this right to science, but it likely includes a continuum of access rights including the right of researchers to access data [17–21]. In the context of data intensive health research, it can also build on the jurisprudence of other human rights such as those to health, to procedural fairness, anti-discrimination, equitable access, and privacy.

In 2014, the Global Alliance for Genomics and Health (GA4GH) began to address one possible aspect of the right to science. Indeed, its *Framework for Responsible Sharing of Genomics and Health-Related Data* [22] as well as accompanying Policies on Consent [23], Privacy and Security [24], Accountability [25], Ethics Review Recognition [26] and Data Sharing Lexicon [27] are centered on the further elaboration of this right in the context of data intensive science.

Focusing on modern, data intensive research, the GA4GH seeks to facilitate global data sharing through the building of enabling policy, IT, and clinical tools. To do this, the policies and tools built by members address difficult issues such as the sharing of legacy data [24], and the need to take a more realistic approach to the evaluation of the actual risk-benefit ratio in international data intensive research as compared to the risks found in interventionist clinical trials. This means weighing the benefits of sharing against concrete, empirically established privacy risks given the nature, sensitivity and level of identifiability of data, rather than treating privacy as absolute, or giving undue weight to hypothetical privacy risks [25]. Moreover, the GA4GH Accountability Policy asks all stakeholders such as researchers, publishers, funders, universities, companies or patient groups that hold data consented for research: "Why are you not sharing?" Or, more importantly, "why are you not designing governance and consent to allow for data sharing?" GA4GH is fostering and building open variant cancer databases (e.g. brcaexchange.org), the matching of individuals and families with rare diseases through "Matchmaker" (matchmakerexchange.org), discovery science through "beacons" using API's (beacon-network.org) and public education and engagement in governance through surveys like YourDNAYourSay.org. The underlying goal is to move away from an automatic ethical presumption that data intensive research is harmful to participants to one that begins with a vision of its potential for the common good as an expression of the fundamental human right to benefit from science.

The GA4GH *Framework* is in the spirit of the opportunity offered by the EU's GDPR that allows organizations to develop Codes of Conduct that can both serve to protect and share data across borders. If approved by the Data Protection Board of the EU Data Commission, such Codes could allow for trans-border data sharing by those promising to adhere thereto. Codes, whether developed by industry, professional bodies or regulators, hold great promise to be flexible instruments for the