Is health-care data the new blood?

Last year, an article published in *The Economist* declared that the world's most valuable resource was no longer oil but data.[1] The piece describes the emergence of a highly lucrative data economy and warns that new antitrust legislation might be needed for greater governance of data. Although many have criticised the comparison of data to oil, specifically surrounding the finite availability of oil as a resource compared with data, the core issue of a highly profitable and under-regulated data economy is real, especially in health care. The health-care data economy is booming with hundreds of start-up companies looking to supposedly fix health care through innovative data, data tools, and technology products. In addition to these legitimate businesses, there is an equally booming shadow economy driven by conventional wisdom that estimates the value of a medical record to be ten times the value of a credit card.[2]

So, in health care, is data equivalent to oil or would it be more accurate to describe it as blood? We propose that health-care data records are digital specimens and should be treated with the same rigour, care, and caution afforded to physical medical specimens. We advocate that the use of these digital samples be limited to validated and beneficial uses for the donor and that patient privacy be fully protected.

Over the past 6 months, high-profile stories and events have highlighted the need to develop more detailed privacy protections and proper usage validations for connected or digital medical technologies. In one case, a continuous proximal airway pressure (CPAP) device manufacturer was sharing patient compliance data from these machines with insurers, who were subsequently denying patient claims on the basis of supposed adherence gaps.[3] In this case, a patient was denied coverage for accessories to the medical device because the device was transmitting usage data to the manufacturer without patient knowledge or consent. This event has raised several extremely important questions. How can patient privacy be defined and protected on connected medical equipment and what are the associated rights of that patient? Are manufacturers obligated to disclose all data being collected and its usage? Are the data collected adequate and properly validated for the intended uses? What consumer protections exist to protect patients in the event of potential discrimination or data misuse?

The rapid pace of connected medical products has regulators and policy experts struggling to understand this extremely diverse and technically complex landscape.[4] Novel applications of technology such as real-time wearable sensors are creating new big data streams that can uniquely identify and physically locate users.[5] Although these technologies pose important privacy and security concerns, in the premarket stages they are subject to the protections of biomedical products premarket regulations for patient protection, such as ethical informed consent and Institutional Review Board oversight as required by the US Food and Drug Administration and the Medical Device Directive in the EU. However, the basic protections of ethical research conduct do not necessarily apply to mature postmarket products. For example, with respect to internet-connected CPAP machines, the data being transmitted to the manufacturer might not be subject to the Healthcare Insurance Portability and Accountability Act (HIPAA) because the data might not contain the explicitly prohibited identifiers or because the manufacturer does not meet the covered entity definition. Under the General Data Protection Regulations (GDPR), the data might be protected as special categories of personal data, but this remains to be tested.

Without appropriate oversight, data quality cannot be guaranteed for unintended uses, an issue that is compounded by the rise of health-care data brokers who have been partnering with the health insurance industry to collect digital specimens on hundreds of millions of Americans.[6] Digital specimens can be medical records, sensor data, race, education level, posts on social media, bill payments, and Amazon orders.[7] According to HIPAA, almost none of these are considered covered entities nor are they subject to the governance or principles of ethical research on humans.

Technology companies rely on contracts such as end-user license agreements (EULAs) and privacy policies to govern the rights to monitor, analyse, and share user data. In instances where a company is not covered by HIPPA, the EULA and other consumer agreements become the primary privacy constraint from a legal perspective. These agreements form the basis of a new-age social contract for how a medical device company would handle a user's digital specimen. However,

today most of the burden of consent resides with the consumer, who is expected to read and understand these privacy policies before using the product, although most people do not. One influential study showed that 97% of users agreed to the privacy policy of a fictitious social network and spent an average of about 70 s to skim the policy, which would normally take about 30 min to read.[8] Today, the contracts are written more to protect companies from lawsuits rather than to establish a set of norms and values around how to handle patient data. This burden should shift more toward the technology company to develop an understandable social contract for the user that clearly outlines how their body-generated data would be used, aggregated, and shared.

Furthermore, although substantial law already provides protections from discrimination caused by genetic data, no such law exists for all these new digital health data streams, and medical device use is far more prevalent in the US population than is genetic testing. The disparity in exposure and risk is extensive. Although the combined US genetic testing market—prenatal or neonate testing and digital genome—is expected to reach US$22 billion by 2024, the medical device industry, which was already $172 billion in 2013, is roughly eight times larger and estimated to account for 4–6% of all US health-care spending.[9,10] Most importantly, a study has shown that consumers are poorly aware of the protections of genetic antidiscrimination law and highly concerned about the effects that optional medical testing might have on their insurability.[11] Clearly, the CPAP incident shows that their concerns are valid.

As complex as these issues are, we propose a three-pronged strategy for avoiding harm and protecting the privacy of digital specimens. First, to enable regulation and protection, digital specimens must be properly categorised by at least three attributes: by data type or format; by level of permission such as consented, unconsented, informed but not consented; and by level of risk to the data donor. Practically, this could be implemented in a similar fashion to the special categories of data within GDPR. Implementation could help ensure that data is validated for quality and accuracy to avoid irresponsible, negligent, or methodologically invalid applications. Second, enabled by this categorisation, new and more practically usable methods of consumer notification must replace or enhance the currently failing End User License Agreement model (also known as the "agree to all the terms listed or you can't use this product" model). Third, consumer protections must be put in place to inform and protect the public but also to enable adequate penalties for privacy violations. In truth, we believe that these steps are the bare minimum that must be accomplished to include, engage, and protect digital specimen donors.

For **provisions of data within GDPR** see https://gdpr-info.eu/art-9-gdpr/

Eric Perakslis, Andrea Coravos
Duke University, Durham, NC, USA (EP); Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA (EP); Elektra Labs, Boston, MA, USA (AC); Harvard-Massachusetts Institute of Technology Center for Regulatory Science, Boston, MA, USA (AC); The Digital Medicine Society, New York, NY, USA (AC)
eperakslis@gmail.com

EP declares no competing interests. AC is a paid employee and shareholder of Elektra Labs.

1 Leaders. Regulating the internet giants: the world's most valuable resource is no longer oil, but data. London: The Economist, May 6, 2017. https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data (accessed Nov 18, 2018).
2 Humer C, Finkle J. Your medical record is worth more to hackers than your credit card. New York: Reuters, Sept 24, 2014. https://www.reuters.com/article/us-cybersecurity-hospitals/your-medical-record-is-worth-more-to-hackers-than-your-credit-card-idUSKCN0HJ21I20140924 (accessed Nov 18, 2018).
3 Allen M. You snooze, you lose: how insurers dodge the cost of popular sleep apnea devices. Washington, DC: National Public Radio, Nov 21, 2018. https://www.npr.org/sections/healthshots/2018/11/21/669751038/you-snoozeyou-lose-how-insurers-dodge-the-costs-of-popularsleep-apnea-devices (accessed Dec 8, 2018).
4 Matwyshyn AM. The 'Internet of Bodies' is here. Are courts and regulators ready? New York: Wall Street Journal, Nov 12, 2018. https://www.wsj.com/articles/the-internet-of-bodies-is-here-are-courts-and-regulators-ready-1542039566 (accessed Jan 3, 2019).
5 Perakslis ED. Protecting patient privacy and security while exploiting the utility of next generation digital health wearables. BMJ Opinion, Jan 18, 2019. https://blogs.bmj.com/bmj/2019/01/18/protecting-patient-privacy-and-security-while-exploiting-the-utility-of-next-generation-digital-health-wearables/ (accessed Jan 18, 2019).
6 Allen M. Pro Publica. Health Insurers are Vacuuming Up Details About You - And It Could Raise Your Rates. July 17, 2018. https://www.propublica.org/article/health-insurers-are-vacuuming-up-details-about-you-and-itcould-raise-your-rates (accessed Jan 17, 2019).
7 Ravindranath M. Does your doctor need to know what you buy on Amazon? Politico, Oct 30, 2018. https://www.politico.com/story/2018/10/30/the-doctor-will-see-through-you-now-893437 (accessed Nov 18, 2018).
8 Obar JA, Oeldorf-Hirsch A. The biggest lie on the internet: ignoring the privacy policies and terms of service policies of social networking services. *Information Commun Soc* 2016; **2016:** 1–20.
9 Global Market Insights. Genetic testing market worth over $22 billion by 2024. Selbyville, DE: Global Market Insights, June 5, 2018. https://globenewswire.com/news-release/2018/06/05/1516735/0/en/Genetic-TestingMarket-worth-over-22-Billion-By-2024-Global-Market-Insights-Inc.html (accessed Nov 18, 2018).
10 MedPac. Report to congress: an overview of the medical device industry. Washington, DC: Medicare Payment Advisory Commission, June, 2017: pp 1–38. http://www.medpac.gov/docs/default-source/reports/jun17_ch7.pdf?sfvrsn=0 (accessed Feb 6, 2019).
11 Parkman AA, Foland J, Anderson B, et al. Public awareness of genetic nondiscrimination laws in four states and perceived importance of life insurance protection. *J Genet Counsel* 2015; **24:** 512–21.