

# COMMENT



**ENVIRONMENT** Will climate change mean more vast algal blooms? **p.349**

**BIOTECHNOLOGY** Lots of ways to alter life but not enough bioethics **p.352**

**TECHNOLOGY** Biography of Marconi explores global communications **p.354**

**CHEMISTRY** Goals change the breath composition of spectators **p.355**

STEPHEN LAM/GETTY



Wearable devices that track fitness are a rich source of real-time health data.

## Stop the privatization of health data

Tech giants moving into health may widen inequalities and harm research, unless people can access and share their data, warn **John T. Wilbanks** and **Eric J. Topol**.

Over the past year, technology titans including Google, Apple, Microsoft and IBM have been hiring leaders in biomedical research to bolster their efforts to change medicine.

In September 2015, Tom Insel announced that he would quit his position as head of

the US National Institute of Mental Health to join Google Life Sciences (now Verily). Three months later, Michael McConnell took a leave of absence from directing major cardiovascular research programmes at California's Stanford University to join him. And last month, Stephen Friend took a

senior position with Apple. He is co-founder and former president of Sage Bionetworks, a non-profit organization that promotes open science and patient engagement in research (where one of us, J.T.W, works).

In many ways, the migration of clinical scientists into technology corporations ►

▶ that are focused on gathering, analysing and storing information is long overdue. Because of the costs and difficulties of obtaining data about health and disease, scientists conducting clinical or population studies have rarely been able to track sufficient numbers of patients closely enough to make anything other than coarse predictions. Given such limitations, who wouldn't want access to Internet-scale, multidimensional health data; teams of engineers who can build sensors for data collection and algorithms for analysis; and the resources to conduct projects at scales and speeds unthinkable in the public sector?

Yet there is a major downside to monoliths such as Google or smaller companies such as consumer-genetics firm 23andMe owning health data — or indeed, controlling the tools and methods used to match people's digital health profiles to specific services.

Digital profiling in other contexts is already creating what has been termed a 'black box' society. Online adverts are tailored to people's age, location, spending and browsing habits. Certain retail services have preferentially been made available only to particular groups of people. And law enforcers are being given tools to help them make sentencing decisions that cannot be openly assessed (see [go.nature.com/29umpu1](http://go.nature.com/29umpu1)). This is all thanks to the deliberately hidden collection and manipulation of personal data.

If undisclosed algorithmic decision-making starts to incorporate health data, the ability of black-box calculations to accentuate pre-existing biases in society could greatly increase. Crucially, if the citizens being profiled are not given their data and allowed to share the information with others, they will not know about incorrect or discriminatory health actions — much less be

able to challenge them. And most researchers won't have access to such health data either, or to the insights gleaned from them.

### INFORMATION FLOW

We believe that the influx of health experts into Silicon Valley could foreshadow a fundamental shift in biomedical research and health care.

In various countries, including the United Kingdom, Sweden, Norway and Estonia, sustained pushes by governments and civil society have made standardized electronic health records the norm. And in the United States, hopes are high that truly useful electronic health records (that can be called up no matter which provider a patient goes to) will be rolled out within the next five years. Meanwhile, advances are being made in machine learning, and there is increased availability of digital health data outside the medical system. This type of information is much more amenable to machine-learning approaches than are conventional clinical observations. Furthermore, some of the technology corporations entering health have hundreds of billions of dollars in cash reserves. These factors combined could enable the new players to eclipse, not just join, the old ones.

Until now, obtaining health data has generally depended on highly skilled professionals who record perhaps tens to hundreds of observations in a clinic or hospital ward once or twice a year, and on researchers who painstakingly extract the relevant information from hard-to-obtain, non-standardized medical records. Now, thanks to smartphone apps — such as those created using Apple's software framework ResearchKit, launched last year — and wearable sensors that can detect gait, location, heart rate and even brain activity, analysts can draw on tens of thousands of real-time observations collected from tens of thousands of people every day, even every minute.

People's search behaviour contains health information too. In February 2015, Google partnered with the Mayo Clinic, a non-profit medical practice and research group based in Rochester, Minnesota, to curate health-related facts, such as the most common causes of low back pain. Google's teams are plugging curated and verified health information into the search engine's smart search algorithm, Knowledge Graph, so that users obtain information that is more relevant and supposedly more accurate whenever they type in their symptoms or condition. The service (which for many ailments may replace visits to the doctor) will only enhance Google's — and only Google's — ability to conduct an unprecedented level of information retrieval for health. By the tech



giant's own calculations, of the 40,000 or so searches that are made every second, 2,000 are health-related.

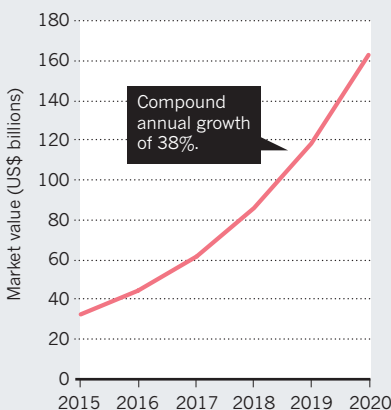
Meanwhile, machine learning is enabling researchers to mine petabytes of data for patterns and associations. The IBM Watson technology platform, for example, sifts through unstructured data using natural language processing and machine learning. Since April last year, IBM has been building 'knowledge experts' for health using the Watson Health cloud platform: customers who store their data in the cloud are given access to algorithms that can help them to make sense of the data. IBM is already using Watson to tackle complex disease problems in collaboration with the University of Texas MD Anderson Cancer Center in Houston — specifically, how to use genomic information from cancer cells to provide individuals with targeted treatments.

Even the rudimentary measurements already flowing from smartphones and wearable sensors can drive better outcomes than decades of medical-device development have been able to provide. For instance, in 2014, a woman with type 1 diabetes wired together a tiny processor, an insulin pump and a continuous glucose monitor to automate the regulation of her blood sugar levels. For a small community of patients, the collective use of such 'home-made' systems has resulted in improvements that are well ahead of those provided

SOURCE: ENMARKETER

### THE DIGITAL HEALTH RUSH

Analysts predict that worldwide, the value of Internet-enabled devices used to collect, analyse and distribute health and fitness data will exceed US\$163 billion within 4 years.







THOMAS PETER/REUTERS

by devices and interventions emerging from conventional markets<sup>1</sup>.

In short, actors big and small are waking up to the enormous profits that could be made from inexpensive consumer health data. Around ten years ago, a scattering of start-ups were obtaining people's genomic information and other health data at low cost and making them available to pharmaceutical partners or other customers. In 2016, every major player in tech — Apple, Google, IBM, Facebook, even Uber — has signalled plans to enter this 'digital health' market.

### A CLOSED LOOP

To be fair, harnessing advances in technology and analytics to radically improve health care is a principal motivator for many corporations. But it is telling that the early players in the game have sequestered information in 'closed loop' systems.

Take the wearable device Enlite, which was made available by the Dublin-based firm Medtronic to people with diabetes in 2013. This stops a pump sending insulin into the wearer's blood when a sensor detects a drop in glucose levels. Although patients can monitor their glucose levels at any instant, their aggregate records are not made accessible to them. And there is no mechanism by which patients or researchers outside the company can gain access to Medtronic's tens of thousands of measurements. The same is true for other wearable devices produced

by the same firm. In fact, the company has refused requests for patients' own heart data (see [go.nature.com/29tdzud](http://go.nature.com/29tdzud)).

Even when corporations do give customers access to their own aggregate data, built-in blocks on sharing make it hard for users to donate them to science. 23andMe, holder of the largest repository of human genomic data in the world, allows users to view and download their own single-letter DNA variants and share their data with certain listed institutions. But for such data to truly empower patients, customers must be able to easily send the information to their health provider, genetic counsellor or any analyst they want.

Pharmaceutical firms have long sequestered limited types of hard-to-obtain data, for instance on how specific chemicals affect certain blood measurements in clinical trials. But they generally lack longitudinal health data about individuals outside the studies that they run, and often cannot connect a participant in one trial to the same participant in another. Many of the new entrants to health, unbound by fragmented electronic health-record platforms, are poised to amass war chests of data and enter them into systems that are already optimized (primarily for advertising) to make predictions about individuals.

**"Of the 40,000 or so searches that are made every second, 2,000 are health-related."**

**A 2011 protest in Berlin called for the right to keep personal digital data private.**

The companies jostling to get into health face some major obstacles, not least the difficulties of gaining regulatory approval for returning actionable information to patients. Yet the market value of Internet-enabled devices that collect and analyse health and fitness data, connect medical devices and streamline patient care and medical research is estimated to exceed US\$163 billion by 2020, as a January report from eMarketer notes (see 'The digital health rush' and [go.nature.com/29fbvch](http://go.nature.com/29fbvch)). Such a tsunami of growth does not lend itself to ethically minded decision-making focused on maximizing the long-term benefits to citizens.

It is already clear that proprietary algorithms can replicate and exacerbate societal biases and structural problems. Despite the best efforts of Google's coders, the job postings that its advertising algorithm serves to female users are less well-paying than are those displayed to male users<sup>2</sup>. A ProPublica investigation in May demonstrated that algorithms being used by US law-enforcement agencies are likely to wrongly predict that black defendants will commit a crime (see [go.nature.com/29aznyw](http://go.nature.com/29aznyw)). And thanks to 'demographically blind' algorithms, in several US cities, black people are about half as likely as white people to live in neighbourhoods that have access to Amazon's one-day delivery service (see [go.nature.com/29kskg3](http://go.nature.com/29kskg3)).

We believe that closed-data and closed-algorithm business models in health — at scale — will hamper scientific progress by blocking the discovery of diverse ways to examine and interpret health data. Longer term, they could increase rather than reduce inequalities and injustices. It is not hard to picture a future in which companies are able to trade people's disease profiles, unbeknownst to the patients. Or one in which health decisions are abstruse and difficult to challenge, and advances in understanding are used to aggressively market health-related services to people — regardless of whether those services actually benefit their health.

### FOR BETTER, NOT WORSE

It is not our intention to demonize technology companies. Closed systems have some innate short-term advantages over open ones. For instance, they can achieve coherence and scale faster than open systems can because their chief executives can dictate formats, standards and norms. But private capital will better serve public interests if at least some layers of the emerging health-research infrastructure are open (the raw data being one layer, the tools for analysis

another and the development of treatments and services for patients a third).

In our view, sensor and other data and the methods for converting those data into clinically usable information should be a public good. The technical and business services that translate insights into value for individual citizens could well be closed. This would be analogous to Google, Apple Maps and Waze providing maps on the back of government-funded geospatial data. Under this scenario, private capital would fund the creation of meaningful patient experiences, not the creation of trade secrets about health.

In theory, hard government regulation could prevent the collection and analysis of digital health data from becoming a high-profit business. The European Union has already passed sweeping directives to protect people's digital information from being exploited for commercial or other purposes. But in the United States at least, the odds are slim that well-constructed digital civil-rights laws will be passed in the next 2–3 years.

Many of the largest tech corporations have come to resemble small nations in their own right: they have enormous 'natural resources' (data and computing power) and global interests to pursue and protect. Just five US-based tech firms — Apple, Microsoft, Alphabet (Google's parent company), Cisco Systems and Oracle — had combined cash reserves of \$504 billion in late 2015, much of which is held offshore to avoid taxation and regulation (see [go.nature.com/29ep3zx](http://go.nature.com/29ep3zx)). Even if the US government wanted to intervene, technologies and their accompanying business models are evolving faster than it can keep up.

In our view, the creation of credible competitors that are open source is the most promising way to regulate closed business models. During the late 1990s, IBM, then one of the biggest players in the software market, abandoned its proprietary web server software in favour of selling services around open-source software. Once users realized that the open-source Apache web server and Linux operating system offered a viable alternative to commercial packages, IBM started selling them its support and configuration services. At around the same time, the release of Netscape's browser source code rapidly fuelled innovation in the browser market and prevented Microsoft from creating a monopoly with Internet Explorer. In both cases, communities of hundreds, or at most a few thousand, transformed the landscape for the world's largest corporations.

The creation of public resources through government funding has a role in this. Soon after the publicly funded Human Genome Project announced in 2000 that a draft was complete, the private sequencing company

Celera stopped charging researchers for access to its data, deposited those data into the public database GenBank and focused instead on trying to develop treatments for disease. Yet the sums of money being directed towards public projects today — such as the Obama administration's Precision Medicine Initiative, which aims to match treatments to patients' genetic and physiological data — pale in comparison to the investment that many companies can bring. (23andMe's latest fundraising round of \$115 million represents 70% of the entire federal investment allotted to the Precision Medicine Initiative.)

What is needed are networks of open projects, combined with sufficient numbers of patients and citizens who are motivated to feed such projects with their health data.

### CASE STUDY

At Sage Bionetworks, we are conducting four ongoing clinical studies on various diseases including Parkinson's. We are also providing the online interfaces for patients to upload their information, and portals to enable data sharing for a further three studies (on diabetes, asthma and cardiac health). Across these seven studies, we have enrolled more than 90,000 participants since March 2015 (see [go.nature.com/29qqk57](http://go.nature.com/29qqk57)). Participants have the automatic right to access and download complete copies of their own data. They also have the right to donate their data for broad reuse by 'qualified researchers', meaning those that have validated their identity, passed a short test and signed a code of ethical practice.

In our studies, more than 75% of participants elect to share their data — presumably to maximize the chances of investigators finding a way to help them and others like them. Our numbers are biased by the fact that we are surveying people who are already enrolled in clinical studies. But in a 2015 survey of a more generic US population by National Public Radio, 53% of those polled said that they would be willing to share their data anonymously with health-care professionals (see [go.nature.com/29qr0b7](http://go.nature.com/29qr0b7)). It could be transformative if even just 5% of the US population donated a copy of their health data to science. After all, it took only a small community of open-source software programmers worldwide to drive major shifts in the computing industry.

In short, a movement not dissimilar to the environmental one is needed to break open how people's data are being used, and to illuminate how they could be used in

*“When it comes to control over our own data, health data must be where we draw the line.”*

the future. In the United States, it was the unified lobbying of small groups of activists in the 1960s and 1970s — each with different reasons for being concerned — that led to a series of groundbreaking federal initiatives, such as the Clean Air Act and the creation of the Environmental Protection Agency. Likewise, at first, data donors may predominantly include those with personal incentives, or those who are philosophically driven to share their health data, for instance through being part of the Quantified Self community, which aims to use technology to measure all aspects of our daily lives. Such early advocates for sharing could help to change norms by pushing for clearer messaging around consent and by raising awareness about what is at stake.

Openness is not an easy goal. Numerous interlocking systems need to be designed, including those to protect privacy, to mitigate harm caused by certain insights becoming public knowledge, and to enable people who do not wish to be phenotyped to opt out. But if the new era of digital health is accompanied by citizen-led pushes for more openness, it may not just be health care that is transformed.

Citizens worldwide have too long a history of being passive players in health care — blindly following instructions from providers. And studies that have tracked reactions to revelations about global surveillance programmes suggest that most people are resigned to the idea that ownership and control of personal information is incompatible with the Internet age<sup>3</sup>.

Yet just as social networking has rocketed around the world in a decade, a worldwide knowledge resource could soon be used to identify the best course of treatment for an individual on the basis of the experiences of millions. This resource will never be built unless each of us takes responsibility for our own health and disease, and for the information that we can generate about ourselves. When it comes to control over our own data, health data must be where we draw the line. ■

**John T. Wilbanks** is chief commons officer at Sage Bionetworks in Seattle, Washington, USA. **Eric J. Topol** is professor of genomics at the Scripps Research Institute in La Jolla, California, USA.  
e-mails: [wilbanks@gmail.com](mailto:wilbanks@gmail.com); [etopol@scripps.edu](mailto:etopol@scripps.edu)

- Lewis, D. M. & Leibrand, S. M. Real-World Use of Open Source Artificial Pancreas Systems. *Am. Diabetes Assoc. 76th Sci. Sessions* (2016); available at <http://go.nature.com/29kiyne>
- Datta, A., Tschantz, M. C. & Datta, A. *Proc. Priv. Enhancing Technol.* **2015**, 92–112 (2015).
- Turow, J., Hennessy, M. & Draper, N. *The Tradeoff Fallacy: How Marketers Are Misrepresenting American Consumers and Opening Them Up to Exploitation* (Univ. Pennsylvania, 2015); available at <http://go.nature.com/29khjju>



### **CORRECTION**

The Comment article 'Stop the privatization of health data' (J. T. Wilbanks & E. J. Topol *Nature* **535**, 345–348; 2016) wrongly stated that the Enlite device sends insulin into the blood when it detects a drop in glucose; in fact, it stops a pump releasing insulin. And 23andMe's latest fundraising round was US\$115 million, not \$150 million.