

# Health Affairs

---

At the Intersection of Health, Health Care and Policy

Cite this article as:

Axel Heitmueller, Sarah Henderson, Will Warburton, Ahmed Elmagarmid, Alex  
"Sandy" Pentland and Ara Darzi  
Developing Public Policy To Advance The Use Of Big Data In Health Care  
*Health Affairs* 33, no.9 (2014):1523-1530  
doi: 10.1377/hlthaff.2014.0771

The online version of this article, along with updated information and services, is  
available at:

<http://content.healthaffairs.org/content/33/9/1523>

**For Reprints, Links &  
Permissions :**

[http://content.healthaffairs.org/1340\\_reprints.php](http://content.healthaffairs.org/1340_reprints.php)

**Email Alertings :** <http://content.healthaffairs.org/subscriptions/etoc.dtl>

**To Subscribe :** <https://fulfillment.healthaffairs.org>

*Health Affairs* is published monthly by Project HOPE at 7500 Old Georgetown Road, Suite 600, Bethesda, MD 20814-6133. Copyright © by Project HOPE - The People-to-People Health Foundation. As provided by United States copyright law (Title 17, U.S. Code), no part of may be reproduced, displayed, or transmitted in any form or by any means, electronic or mechanical, including photocopying or by information storage or retrieval systems, without prior written permission from the Publisher. All rights reserved.

Not for commercial use or unauthorized distribution

By Axel Heitmueller, Sarah Henderson, Will Warburton, Ahmed Elmagarmid, Alex “Sandy” Pentland, and Ara Darzi

DOI: 10.1377/hlthaff.2014.0771  
HEALTH AFFAIRS 33,  
NO. 9 (2014): 1523–1530  
©2014 Project HOPE—  
The People-to-People Health  
Foundation, Inc.

## ANALYSIS &amp; COMMENTARY

# Developing Public Policy To Advance The Use Of Big Data In Health Care

**ABSTRACT** The vast amount of health data generated and stored around the world each day offers significant opportunities for advances such as the real-time tracking of diseases, predicting disease outbreaks, and developing health care that is truly personalized. However, capturing, analyzing, and sharing health data is difficult, expensive, and controversial. This article explores four central questions that policy makers should consider when developing public policy for the use of “big data” in health care. We discuss what aspects of big data are most relevant for health care and present a taxonomy of data types and levels of access. We suggest that successful policies require clear objectives and provide examples, discuss barriers to achieving policy objectives based on a recent policy experiment in the United Kingdom, and propose levers that policy makers should consider using to advance data sharing. We argue that the case for data sharing can be won only by providing real-life examples of the ways in which it can improve health care.

In going about their daily lives, people leave behind huge quantities of “digital breadcrumbs”—the small, often unconnected pieces of information that are produced through the use of cell phones and the Internet as well as other trackable daily activities such as commuting or shopping. This vast ocean of data gives clues to people’s behavior, preferences, and future actions. The data also offer significant opportunities for researchers, clinicians, and policy makers to move away from looking at population averages and toward the use of personalized information that has great potential to generate personal, societal, and commercial benefits. These opportunities are particularly significant in health care. They include the real-time tracking of outbreaks of diseases, predictions of future outbreaks, and the development of personalized medicine.<sup>1</sup>

Despite the great number of potential benefits,

though, formidable challenges remain. Arguably, health information is materially different from other forms of data because of the importance that people place on their privacy. The unintended release of a person’s health information into the public realm has huge potential to undermine personal dignity and cause embarrassment and financial harm. Without strong privacy safeguards, patients are less likely to take part in research and more likely to engage in protective behaviors such as withholding information from health care professionals, moving from one doctor to another, or avoiding the health care system completely.<sup>2</sup>

Much attention has been paid to specific aspects of big data (see our discussion of the definition of *big data* below) and the opportunities for improving the quality and value of health care that are available. Less effort has been made to structure the debate and support decision makers in formulating policies to realize these

**Axel Heitmueller** (Axel.Heitmueller@imperialcollegehealthpartners.com) is policy fellow of the Big Data and Health Forum at the World Innovation for Summit for Health (WISH), Qatar Foundation, and an honorary fellow at the Centre for Health Policy, Institute of Global Health Innovation, Imperial College London, in the United Kingdom.

**Sarah Henderson** is head of forum development at WISH and a policy fellow at the Institute of Global Health Innovation, Imperial College London.

**Will Warburton** is forum director at WISH and a senior policy fellow at the Centre for Health Policy, Institute of Global Health Innovation, Imperial College London.

**Ahmed Elmagarmid** is executive director of the Qatar Computing Research Institute, Qatar Foundation, in Doha.

**Alex “Sandy” Pentland** is chair of the Big Data and Health Forum for WISH and director of the Human Dynamics Laboratory, Massachusetts Institute of Technology, in Cambridge, Massachusetts.

**Ara Darzi** is executive chair of WISH, Qatar Foundation, and director of the Institute of Global Health Innovation, Imperial College London.

opportunities and frame the key policy trade-offs.

This article therefore explores four questions that we think are central to the development of effective policies in this area: What aspects of big data are most relevant for health care? What is the policy's intention? What are the barriers to achieving the policy's objectives? What policy levers are available? The article provides a detailed discussion of each question to help structure policy makers' thinking.

### What Aspects Of Big Data Are Most Relevant For Health Care?

Multiple concepts are nested inside the term *big data*, for which there is no single generally accepted definition. The term most commonly refers to data characterized by great volume, variety, and velocity.<sup>3</sup> Instead of adding to the many existing definitions, we offer a simple taxonomy of four data types that is framed in terms of who controls access to those types and that builds on previous work<sup>4</sup> (Exhibit 1).

**PERSONAL AND PROPRIETARY DATA** The first data type is personal and proprietary data, which are controlled by individuals, commercial entities, or nongovernment organizations. Typically the individual or commercial entity has the desire, and often the legal right, to restrict access to and use of the data. Consequently, there must be

both legal and technological infrastructures that can strictly control and audit that use.

Examples of these data are information on people's shopping behavior or personal health records. Perhaps more important, this type of data also includes the digital breadcrumbs that people produce, including cell phone and credit card information. Commercial interests and privacy concerns are clearly relevant here, and we will explore them more fully below.

The level of access to these data may vary. For example, private-sector organizations such as cell phone providers may make call data available for public-interest initiatives.<sup>5</sup>

**GOVERNMENT-CONTROLLED DATA** The second type is data to which a government can restrict access. These data include personal tax and census data and, in some countries, personal health records. The advent of big data in health care may dramatically expand the depth and breadth of these data to include data on individual behavior. These could provide an important source of information for policy design. However, data breaches during the past decade, such as those by the US National Security Agency, highlight the potential risks to governments and the wider society from holding and using such data.

The level of restriction on access to these data may vary. For example, anonymized hospital episode statistics in the United Kingdom are

#### EXHIBIT 1

##### Types Of Data And Levels Of Access

| Type of data                  | Level of access  |  |   |
|-------------------------------|--|--|---|
|                               | None   | Limited  | Unrestricted  |
| Personal and proprietary data | Personal bank and income information, commercial information | Anonymized cell phone data for research, <sup>a</sup> consumer data offered for a fee <sup>b</sup>   | Twitter and other social media data; raw GPS location data, such as is used for public transport apps; personal health information voluntarily shared on websites such as PatientsLikeMe <sup>c</sup> |
| Government-controlled data    | Personal tax records, classified intelligence                | Anonymized raw health data (for example, available following an assessment of suitability and eligibility of the applicant by a panel of experts), one-off government hackathons, where computer programmers engage in collaborative data work to find solutions to particular challenges <sup>d</sup> | Information about civil servants' pay in some countries, such as the United Kingdom; raw government accounts data <sup>e</sup>  |
| Open data commons             | — <sup>f</sup>   | Survey and census data (may require users to register); <sup>g</sup> UK Land Registry information (requires users to register and pay a fee)   | Maps, weather data, traffic information, financial information about companies  |

**SOURCE** Authors' analysis of data from the sources listed in the exhibit footnotes. <sup>a</sup>Poole J. Winning research from the Data 4 Development challenge (Note 5 in text). <sup>b</sup>For example, see Press release, Johnson & Johnson announces clinical trial data sharing agreement with Yale School of Medicine (Note 33 in text). <sup>c</sup>PatientsLikeMe (Note 39 in text). <sup>d</sup>For examples, see the following sources: (1) Open City (Note 9 in text). (2) Tech City UK convenes developer talent and tech community to host #floodhack in Shoreditch (Note 10 in text). <sup>e</sup>Gov.UK. Guidance: whole of government accounts 2013 to 2014: guidance for preparers [Internet]. London: Whole of Government Accounts; [updated 2014 May 29; cited 2014 Jul 31]. Available from: <https://www.gov.uk/government/publications/whole-of-government-accounts-2013-to-2014-guidance-for-preparers>. <sup>f</sup>Not applicable. <sup>g</sup>For example, UK Data Service [home page on the Internet]. Colchester (UK): The Service; [cited 2014 Jul 31]. Available from: <http://ukdataservice.ac.uk/>.

government-controlled data but may be made available for third-party use for research purposes, subject to a vetting process.

**OPEN DATA COMMONS** The third data type is open data commons. These data are available to all, with at most minor obligations such as acknowledging the source of the data when they are used. The data may be private, commercial, or government controlled. In contrast to open data, open data commons are usually kept up-to-date and provided in accessible format—enabling the immediate use of information.<sup>6</sup>

A key presumption is that data are worth more when they are shared: Previously unconnected data such as personal health records from different health and social care providers can be linked and analyzed to provide new insights that can inform improvements in sectors of the economy such as health care, transportation, and government. Therefore, using an open data commons may provide unprecedented opportunities to assess the efficacy of public policies.

In many countries, various data commons exist for geographic, climate, census, and financial data. With the advent of big data, there exists the potential to develop many more types of data commons. These commons can be both real-time and unprecedented in their level of detail because they can rely on data already produced as a side effect of daily life, such as digital transaction records and cell phone location fixes.

One major concern with such data commons is that they could endanger personal privacy. As mentioned above, this is particularly important when considering health information because of potential adverse behavioral changes.

According to Demos, a UK think tank, “there is a ‘crisis of confidence’ in the way that personal information and behaviour data are being used.”<sup>7</sup> Interestingly, “85 per cent of the public use store loyalty cards, despite these worries.”<sup>7</sup> These results suggest that views about personal health information are more complex than views about other data.

### What Is The Policy’s Intention?

Successful public policy requires clarity of objective.<sup>8</sup> Given the multiple dimensions of the big-data debate, being clear about a policy’s intention is key.

To illustrate this point, we can draw on the above taxonomy to identify two objectives that policy makers may want to pursue. Achieving these would begin to maximize the benefits of big data. Of course, they are not the only relevant objectives.

One objective that policy makers could pursue is to link personal, proprietary, and government

data, both within and across the three categories described above, to pursue health care objectives. In most health care systems, patients’ records are not linked across providers to enable the analysis of patient pathways (this is particularly true for private health care systems). Focusing on linking data could improve care planning at both the patient and population levels. Arguably, there is even more potential in linking data across sectors—for example, the retail and health sectors—which could aid targeted public health campaigns.

Another objective that policy makers could pursue would be to increase the amount of information that is available as open data commons. This includes both the expansion of the amount of data available and also, where acceptable, making sensitive data (following anonymization or the use of other privacy safeguards) more broadly accessible.

The real power of open data is the unexpected uses or applications that could arise. Even the most farsighted policy maker will not be able to predict the multiplicity of applications that are likely to ensue if data are made available publicly. Events such as “hack days”—when programmers, analysts, and other stakeholders come together to collaborate intensively on a project—have led to interesting policy solutions, and the more data that are publicly available, the greater the opportunities.<sup>9,10</sup>

### What Are The Barriers To Achieving The Policy’s Objectives?

Once the policy intention has been defined, understanding the barriers to achieving the objective becomes paramount. In this section we draw on the literature and a recent policy experiment in the United Kingdom to provide a discussion of such barriers.

The United Kingdom’s National Health Service (NHS) celebrated its sixty-fifth birthday in 2013 and is one of the most cherished public institutions in the country. Nigel Lawson, former chancellor of the exchequer, famously called the NHS “the closest thing the English have to a religion.”<sup>11</sup>

Because the NHS is a national system, it provides significant opportunities to work with population data across community, mental health, primary, secondary, and tertiary services. However, health records are held at the provider level in accordance with the existing data protection rules, and there is no national repository for health data.<sup>12</sup> There has been some progress at the local level, but broader attempts to link data across the NHS have been unsuccessful.

The most pertinent example of the difficulty of

linking health information is the NHS's Connecting for Health project, which was launched in 2002. Its prime focus was to move toward a single electronic health record that would connect primary care providers and roughly three hundred hospitals. The project was eventually abandoned after around £12 billion (\$20 billion) had been spent on it, after it attracted strong criticism from the government's Public Accounts Committee.<sup>13</sup>

The latest health records initiative is a program known as *care.data*,<sup>14</sup> which aims to link, for the first time, the vast majority of health records in England. If successful, it would create one of the world's largest health care databases. However, the program has run up against significant opposition, particularly from primary care providers and patient advocacy groups.<sup>15</sup> This opposition provides recent and practical insights into the challenges faced by policy makers.

We distinguish between three broad categories of barriers: normative barriers, market failures, and technocratic barriers—that is, barriers related to technological issues and to government processes and rules.

**NORMATIVE BARRIERS** This category includes cultural and ethical norms. These norms affect how trusting people are that the confidentiality of their data will be protected. And at least in the intermediate term, the norms will set boundaries on the use of big data.<sup>16</sup>

For example, most patients are likely to accept the need for sharing health records across providers for the purpose of improving service quality and planning. In contrast, the idea that *care.data* would be available to third parties (specifically those in the private sector), even if anonymized, proved to be controversial.<sup>17</sup> The debate about the ethical implications of applying predictive tools to stratify data and target care at specific subpopulations shows that even applications restricted to health care can raise concerns.<sup>18</sup>

These concerns are partly the consequence of the absence of an ethical framework that would allow a clear demarcation of boundaries and expression of public preferences. These boundaries and preferences vary greatly across countries.<sup>19</sup> For example, a recent study found that people in India are more than four times more willing than people in Sweden to share personal information online to receive personalized services.<sup>20</sup>

**MARKET FAILURES** A key assumption in the big-data debate, particularly in the health care context, is that broad societal benefits can be achieved from sharing and linking data. These “positive externalities” are often said to be in conflict with the private interests of those who

own or control the data. If these individuals or organizations cannot internalize the broad benefits, they are unlikely to have either monetary or nonmonetary incentives to share their data.

To illustrate this point and show how it interlinks with social norms, consider the following example. Supermarkets routinely collect data from their customers, mostly through loyalty and credit cards.<sup>21,22</sup> The volume and velocity of transactions allow the creation of algorithms to target customers with personalized marketing. Sharing such data with health authorities could help personalize efforts to promote public health. However, this would require customers and supermarkets to recognize the societal benefits of sharing data and believe that they outweigh privacy and commercial concerns, respectively.

Different policy responses are therefore required, depending on the position that societies take on the individual-collective benefit continuum. In a highly private society, for example, it may be possible for retailers to offer personalized health assessments for a fee to customers. This would internalize benefits for the supermarket and would also mean that personal customer information was not shared. At the other extreme, retailers may be required to share customer information as a matter of course. This may improve preventive health care. However, it may also hamper the speed of innovation and undermine public trust, with adverse behavioral consequences.<sup>2</sup>

There are a number of intermediate options. These include allowing customers to opt in or out of data-sharing programs and employing safeguards such as the anonymization of information.

In the case of *care.data*, one of the externalities concerns the use of anonymized data by third parties to conduct product and clinical research. The availability of longitudinal data that contain detailed care records for tens of millions of patients is seen as transformative in developing and monitoring the efficacy of drugs, for example.<sup>23</sup> This may lead to both commercial benefits for industry and improved treatment options. However, the real or perceived risks to privacy at the individual level may outweigh these wider benefits.

**TECHNOCRATIC BARRIERS** Technocratic barriers include those related to technology, common data standards, interoperability across data systems, and—perhaps most important—legislation. The capability of governments to safely process and store large sets of sensitive personal or commercial data is also a consideration. Examples of policy measures to overcome barriers are discussed in the next section.



## What Policy Levers Are Available?

Two inherent trade-offs are relevant to the use of big data in health care and public policy making. The first trade-off concerns the role of government in simultaneously protecting people's privacy and taking advantage of the benefits of large data sets—a role that requires careful contextual calibration. The second trade-off is related to the tension between realizing the societal benefits involved in sharing data and safeguarding proprietary rights. We argue that balancing collective interests and individual and organizational interests is a fundamental and universal challenge.

In what follows, we briefly consider a number of areas that policy makers could focus on to help achieve the objectives described above. However, the suitability, feasibility, and acceptability of policy measures will depend on the circumstances of different nations.

**BUILDING SUPPORT FOR DATA SHARING** Successful policy is based on a robust case for change that is understood and strongly supported by those affected. Evidence shows that people are fearful of sharing their data largely because companies and government do not usually clearly explain how they intend to use the information.<sup>1</sup> Making this case effectively requires winning the rational as well as the emotional argument by demonstrating the real-world benefit to (in the case of health care) patients and providers. Employing deliberative democracy approaches—for example, holding events to increase patient engagement—has been shown to have some potential to facilitate public understanding and support a case for change.<sup>24</sup> The failure to clearly explain the reasons for data sharing to patients and, in particular, primary care providers early on has been a significant shortcoming of care .data.

Policy makers in the United Kingdom have been accused of assuming that the benefits of data sharing were self-evident and would not require much explanation. When it became clear that there would be considerable opposition to care.data, especially to the provision of data to third parties, the NHS sent every household in the country a leaflet that described the benefits and offered an opt-out clause.<sup>25</sup> However, this national campaign was not accompanied by a sufficient local program of engagement with patients and the small but vocal group of primary care providers.

Furthermore, the national campaign failed to reveal or confront the potential perverse incentives for primary care providers. For example, greater transparency of performance from shared data would make it easier to compare providers in terms of quality and cost-effective-

ness. Only in 2014 have attempts been made to generate examples of real-world applications<sup>26</sup> and to build a coalition of supporters by piloting the program locally.

At the same time, there have been excellent examples of data sharing that are grounded in real-world applications and that were produced by patients, commissioners, and primary care professionals, which have resulted in benefits to patients and the local health economy.<sup>27</sup> Their distinguishing quality is that they did not start with a discussion about data sharing but with the desire to improve care, and they treated data sharing as one important way to do this.

It will be interesting, in years to come, to compare the relative merits of the United Kingdom's approach and the US government's review of big data, launched in June 2014,<sup>28</sup> particularly in terms of public trust and buy-in.

**BUILDING THE EVIDENCE BASE** Making the case for data sharing requires evidence that changes actually deliver the promised benefits. There is little systematic and robust evidence of how data sharing has resulted in benefits in health care beyond promising case studies.<sup>1</sup> One crucial role for government, then, is to support the systematic gathering of knowledge.

**LEADING BY EXAMPLE** Governments hold vast amounts of data and should lead the debate about increasing access to data by moving as much as possible into open data commons. Many Western countries have made huge advances in this area. For example, the French government recently launched an open data platform.<sup>29</sup> Similar initiatives are being developed in the United Kingdom<sup>30,31</sup> and in the United States.<sup>32</sup> There are also a growing number of examples in developing countries.<sup>33</sup>

In addition, the nongovernmental sector has a role to play in this area. One example is sharing clinical trial data to advance science.<sup>34</sup>

**CREATING DEMAND AND CAPABILITY** In many countries health care is, to a greater or lesser extent, the responsibility of governments. Governments can therefore play an important role in creating demand for data applications, such as personalized and targeted care plans, given the limited but growing private market for these services.

Stimulating and incentivizing innovation outside government bodies by, for example, funding small and medium-size enterprises or creating public-private partnerships between health care providers and industry could be an area of focus for policy makers.<sup>35</sup> This may be achieved through strategies involving more experimentation with and small-scale tests and deployments of data applications that are continuously evaluated and revised. So-called living laboratories in

which entire communities are allowed to experiment with different frameworks of big data could make it possible to explore the risks and rewards more quickly.

Governments can also stimulate innovation in data sharing by incentivizing service-model change. For example, in the United States the Affordable Care Act encourages the development of accountable care organizations (ACOs). The ACO model puts a strong emphasis on payments for outcomes—which in turn incentivizes or requires information sharing between payers and providers—and allows patients access to their records.<sup>36</sup>

Finally, government has a role to play in capacity and capability development. There is an identified shortage of data analysts and appropriately trained clinicians, managers, and practitioners to drive the use of big data.<sup>37,38</sup> In response, the UK government, for example, has announced a £40 million (\$67 million) investment to set up the Alan Turing Institute to drive data algorithm research.<sup>39</sup> The institute will have a broad remit around research, education, and knowledge transfer in the data sciences and will cover more than just health care.

**DEVOLVING RESPONSIBILITY TO PATIENTS** A complementary strategy to sharing data between providers is to make patients the data connectors and give them online access to their records. Programs such as PatientsLikeMe<sup>40</sup> and Patients Know Best<sup>41</sup> are beginning to show both the power that the devolution of responsibility of data ownership to patients can have to improve health care and also how patients decide who they want to share their information with. Enabling patients to control their health care records may also be a viable alternative to the notoriously difficult task of making existing provider systems interoperable.<sup>42</sup>

**CREATING TRUST NETWORKS** In a trust network, all personal data have labels that specify what the data can and cannot be used for. These labels correspond to the terms in a legal contract signed by all of the participants, which states the penalties for not obeying the labels and gives the participants the right to audit the use of the data. Once all of the permissions and information about data origin and ownership are in the system, data use can be automatically audited, and participants can change their permissions or withdraw data.

Some trust networks have been shown to be secure. One example is the SWIFT network, which reliably handles trillions of dollars of interbank transfers per day. The financial and health care contexts cannot be directly compared, but the existence of the SWIFT network shows that trust networks on a large scale can

handle complex information securely: The SWIFT network has never been hacked.<sup>4</sup>

To give people a similarly safe method of managing personal data, researchers have built open-source software systems such as Open mhealth<sup>43</sup> and openPDS.<sup>44</sup> And trust networks for personal data management have been proposed within the World Economic Forum's personal data initiative.<sup>45</sup>

**LEGISLATING SMARTLY** Legislation is both a key enabler of, as well as a barrier to, the use of big data. Most developed countries have some form of data protection and information governance legislation. However, few of these laws were written with big data in mind.<sup>46</sup> Thus, it may be necessary to revisit the question of whether existing legislation is still achieving the right balance between benefits and risks.

In the case of care.data, for example, significant legislative weaknesses in the oversight of how data could be shared with third parties were discovered too late, further undermining the trust of patients and professionals. These weaknesses have now been addressed, and the results include stronger enforcement of penalties following breaches, a simplification from multiple data types to only two (anonymized and non-anonymized), and a legal requirement that the data can be used only for care purposes in a strictly controlled environment (sometimes referred to as a "data lab" or "fume cupboard").<sup>47</sup>

Conflicts between legislative jurisdictions can also present a barrier to the use of big data. For example, a number of changes have been proposed to the European General Data Protection Regulation, including an opt-in requirement for data sharing that is the opposite of the opt-out clause used in care.data.<sup>48</sup>

The challenges for legislators include the speed of change and the uncertainties of future risks and benefits. For example, it is hard to anticipate the full risks and benefits of genomics and biobanks in the years to come. It has been argued that to secure public support, more adoptive legislation and an explicit acknowledgment of the complexity of the issues is required. For example, Mayo Clinic in the United States employed such an approach for its governance around genomic research.<sup>49</sup>

Developed countries may have well-established but out-of-date legislative systems that can act as a barrier to the use of big data. In contrast, the challenge for developing countries is likely to be that only limited appropriate legislation is in place. However, this presents an opportunity to build a basic but progressive information governance system that leapfrogs those more established but less agile systems in developed countries.

## Conclusion

Understanding the potential of big data in health care has only just begun. This article has explored four central questions that policy makers should consider when developing public policy about the use of big data in health care. Recent policy initiatives demonstrate that the issues are complex and constantly evolving. Legislators

have struggled to keep pace and may have underestimated the magnitude of change required to fully realize the benefits of big data. Transparency from the outset with both patients and practitioners about benefits as well as risks will be central to winning the rational as well as the emotional case in favor of increased data sharing. ■

Some of the material in this article was previously presented at the World Innovation Summit for Health, an initiative of Qatar Foundation, in Doha, Qatar, December 11, 2013. The authors acknowledge Qatar Foundation funding for some of the research in this article.

## NOTES

- 1 For more examples of big data applications in health care, see Pentland A, Reid TG, Heibeck T. Revolutionizing medicine and public health: report of the Big Data and Health Working Group 2013 [Internet]. Doha (Qatar): World Innovation Summit for Health; 2013 Dec [cited 2014 Jul 31]. Available from: <http://www.wish-qatar.org/app/media/382>
- 2 Goldman J. Protecting privacy to improve health care. *Health Aff (Millwood)*. 1998;17(6):47–60.
- 3 Dumbill E. What is big data? An introduction to the big data landscape. Radar [serial on the Internet]. 2012 Jan 11 [cited 2014 Jul 31]. Available from: <http://radar.oreilly.com/2012/01/what-is-big-data.html>
- 4 Bilbao-Osorio B, Dutta S, Lanvin B, editors. The global information technology report 2014: rewards and risks of big data [Internet]. Geneva: World Economic Forum; 2014 [cited 2014 Jul 31]. Available from: [http://www3.weforum.org/docs/WEF\\_GlobalInformationTechnology\\_Report\\_2014.pdf](http://www3.weforum.org/docs/WEF_GlobalInformationTechnology_Report_2014.pdf)
- 5 For example, see Poole J. Winning research from the Data 4 Development challenge. United Nations Global Pulse [blog on the Internet]. 2013 May 6 [cited 2014 Jul 31]. Available from: <http://www.unglobalpulse.org/D4D-Winning-Research>
- 6 For a definition and overview of open data, see Manyika J, Chui M, Farrell D, Van Kuiken S, Groves P, Doshi EA. Open data: unlocking innovation and performance with liquid information [Internet]. New York (NY): McKinsey and Company; 2013 Oct [cited 2014 Jul 31]. Available from: [http://www.mckinsey.com/insights/business\\_technology/open\\_data\\_unlocking\\_innovation\\_and\\_performance\\_with\\_liquid\\_information](http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information)
- 7 Bartlett J. The data dialogue [Internet]. London: Demos; 2012 [cited 2014 Jul 31]. Available from: [http://www.demos.co.uk/files/The\\_Data\\_Dialogue.pdf?1347544233](http://www.demos.co.uk/files/The_Data_Dialogue.pdf?1347544233)
- 8 For example, see Hallworth M with Parker S, Rutter J. Policy making in the real world: evidence and analysis [Internet]. London: Institute for Government; 2011 Apr [cited 2014 Aug 1]. Available from: <http://www.instituteforgovernment.org.uk/sites/default/files/publications/Policy%20making%20in%20the%20real%20world.pdf>
- 9 For US examples, see Open City [home page on the Internet]. Chicago (IL): Open City; [cited 2014 Jul 31]. Available from: <http://opencityapps.org/>
- 10 The UK government recently invited participants to a “#floodhack”—that is, an opportunity for the technology community to use otherwise restricted government data—in London to find digital solutions to recent flooding. See Tech City UK convenes developer talent and tech community to host #floodhack in Shore-ditch. Tech City UK Press blog [blog on the Internet]. 2014 Feb 17 [cited 2014 Jul 31]. Available from: <http://techcityukpressoffice.tumblr.com/post/76945685131/tech-city-uk-convenes-developer-talent-and-tech>
- 11 Lawson N. The view from No. 11: memoirs of a Tory radical. London: Bantam Press; 1992. p. 613.
- 12 Data protection in the United Kingdom is governed by the Data Protection Act of 1998. For details, see Information Commissioner's Office. Data protection [Internet]. Wilmslow (UK): ICO; [cited 2014 Aug 1]. Available from: [http://ico.org.uk/for\\_organisations/data\\_protection](http://ico.org.uk/for_organisations/data_protection)
- 13 House of Commons Committee of Public Accounts. The National Programme for IT in the NHS: an update on the delivery of detailed care records systems [Internet]. London: Committee of Public Accounts; 2011 Aug 3 [cited 2014 Aug 1]. Available from: <http://www.publications.parliament.uk/pa/cm201012/cmselect/cmpubacc/1070/1070.pdf>
- 14 National Health Service. NHS choices: your records [Internet]. London: NHS; [last reviewed 2013 Feb 1; cited 2014 Aug 1]. Available from: <http://www.nhs.uk/NHSEngland/thenhs/records/healthrecords/Pages/care-data.aspx>
- 15 Hoeksma J. The NHS's care.data scheme: what are the risks to privacy? *BMJ*. 2014;348:g1547.
- 16 Ensminger J, Knight J. Changing social norms: common property, bridewealth, and clan exogamy. *Curr Anthropol*. 1997;38(1):1–24.
- 17 Sheather J, Brannan S. Patient confidentiality in a time of care.data. *BMJ*. 2013;347:f7042.
- 18 Cohen IG, Amarasingham R, Shah A, Xie B, Lo B. The legal and ethical concerns that arise from using complex predictive analytics in health care. *Health Aff (Millwood)*. 2014;33(7):1139–47.
- 19 EMC. EMC Privacy Index: global results [Internet]. Hopkinton (MA): EMC; [cited 2014 Aug 1]. Available from: <http://www.emc.com/campaign/privacy-index/global.htm>
- 20 Ipsos MORI [Internet]. London: Ipsos MORI. Press release, Three in four Britons are worried about collecting information about them; 2014 Feb 11 [cited 2014 Aug 1]. Available from: <http://www.ipsos-mori.com/research/publications/researcharchive/3342/Three-in-four-Britons-are-worried-about-companies-collecting-information-about-them.aspx>
- 21 For example, see BigData-Startups. Walmart is making big data part of its DNA [Internet]. The Hague: BigData-Startups; 2013 [cited 2014 Aug 1]. Available from: <http://www.bigdata-startups.com/BigData-startup/walmart-making-big-data>



- part-dna/
- 22 Ferguson D. How supermarkets get your data—and what they do with it. *Guardian* [serial on the Internet]. 2013 Jun 7 [cited 2014 Aug 1]. Available from: <http://www.theguardian.com/money/2013/jun/08/supermarkets-get-your-data>
- 23 Vandenbroucke JP, Psaty BM. Benefits and risks of drug treatments: how to combine the best evidence on benefits with the best data about adverse effects. *JAMA*. 2008; 300(20):2417–9.
- 24 McGuire AL, Basford M, Dressler LG, Fullerton SM, Koenig BA, Li R, et al. Ethical and practical challenges of sharing data from genome-wide association studies: the eMERGE Consortium experience. *Genome Res*. 2011;21(7):1001–7.
- 25 National Health Service. Better information means better care [Internet]. London: NHS; [cited 2014 Aug 1]. Available from: <http://www.nhs.uk/NHSEngland/thenhs/records/healthrecords/Documents/better-information-means-better-care.pdf>
- 26 For example, see GovLab. Released: the open data era in health and social care [Internet]. New York (NY): Governance Lab at New York University; [cited 2014 Aug 1]. Available from: <http://thegovlab.org/nhs/>
- 27 For example, patients, commissioners of local health services, and physicians jointly created personalized information for patients with chronic obstructive pulmonary disease in London using big data. See NHS Redbridge. COPD care checklist: putting data in the hands of patients to drive care quality and reduce hospital admissions [Internet]. London: NHS Redbridge; 2012 May [cited 2014 Aug 1]. Available from: <http://www.innovationunit.org/sites/default/files/COPD%20Care%20Checklist%20Report.pdf>
- 28 White House. The 90-day review for big data [Internet]. Washington (DC): White House; [cited 2014 Aug 1]. Available from: <http://www.whitehouse.gov/issues/technology/big-data-review>
- 29 Data.gouv.fr [home page on the Internet]. Paris: Etalab; 2013 [cited 2014 Aug 1]. Available from: <http://www.data.gouv.fr/en/>
- 30 Data.gov.uk [home page on the Internet]. London: UK Government; [cited 2014 Aug 1]. Available from: <http://data.gov.uk>
- 31 Open Data Institute. About the ODI [Internet]. London: ODI; [cited 2014 Aug 1]. Available from: <http://theodi.org/about-us>
- 32 White House. Open Government Initiative [Internet]. Washington (DC): White House; [cited 2014 Aug 1]. Available from: <http://www.whitehouse.gov/open>
- 33 Open Data Research Network. Exploring the emerging impacts of open data in developing countries [Internet]. Geneva: World Wide Web Foundation; [cited 2014 Aug 1]. Available from: <http://www.opendataresearch.org/emergingimpacts>
- 34 For example, see Johnson & Johnson [Internet]. New Brunswick (NJ): Johnson & Johnson. Press release, Johnson & Johnson announces clinical trial data sharing agreement with Yale School of Medicine; [last updated 2014 Jul 31; cited 2014 Jul 31]. Available from: <http://www.jnj.com/news/all/johnson-and-johnson-announces-clinical-trial-data-sharing-agreement-with-yale-school-of-medicine>
- 35 For examples of how industry responds to big data opportunities in health, see Groves P, Kayyali B, Knott D, Van Kuiken S. The “big data” revolution in healthcare: accelerating value and innovation [Internet]. New York (NY): McKinsey and Company; 2013 Jan [cited 2014 Aug 1]. Available from: [http://www.payerfusion.com/wp-content/uploads/2014/02/The\\_big\\_data\\_revolution\\_in\\_healthcare-1.pdf](http://www.payerfusion.com/wp-content/uploads/2014/02/The_big_data_revolution_in_healthcare-1.pdf)
- 36 Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff (Millwood)*. 2014; 33(7):1123–31.
- 37 Gov.UK. UK data capability strategy: seizing the data opportunity [Internet]. London: Department for Business, Innovation, and Skills; [updated 2013 Oct 31; cited 2014 Aug 1]. Available from: <https://www.gov.uk/government/publications/uk-data-capability-strategy>
- 38 Policy Exchange. The big data opportunity: making government faster, smarter, and more personal [Internet]. London: Policy Exchange; 2012 Jul 3 [cited 2014 Aug 1]. Available from: <http://www.policyexchange.org.uk/>
- 39 BBC News. Alan Turing Institute to be set up to research big data. *BBC News* [serial on the Internet]. 2014 Mar 19 [cited 2014 Aug 1]. Available from: <http://www.bbc.co.uk/news/technology-26651179>
- 40 PatientsLikeMe [home page on the Internet]. Cambridge (MA): PatientsLikeMe; [cited 2014 Jul 31]. Available from: <http://www.patientslikeme.com>
- 41 Patients Know Best [home page on the Internet]. [Please provide city] (UK): Patients Know Best; [cited 2014 Aug 1]. Available from: <http://www.patientsknowbest.com>
- 42 Khan JS, Aulakh V, Bosworth A. What it takes: characteristics of the ideal personal health record. *Health Aff (Millwood)*. 2009;28(2):369–76.
- 43 Open mHealth [home page on the Internet]. San Francisco (CA): Open mHealth; [cited 2014 Aug 1]. Available from: <http://openmhealth.org>
- 44 OpenPDS [home page on the Internet]. Cambridge (MA): OpenPDS; [cited 2014 Aug 1]. Available from: <http://openpds.media.mit.edu>
- 45 World Economic Forum. Rethinking personal data [Internet]. Geneva: The Forum; [cited 2014 Aug 1]. Available from: <http://www.weforum.org/issues/rethinking-personal-data>
- 46 Terry NP. Protecting patient privacy in the age of big data. *UMKC Law Rev*. 2012;81(2):385–415.
- 47 Kelsey T. ROCR approval applied for [Internet]. Leeds (UK): NHS England; 2014 Apr 14 [cited 2014 Aug 1]. (Area Team Reference No. 01515). Available from: <http://www.england.nhs.uk/wp-content/uploads/2014/04/cd-stakeholder-lett.pdf>
- 48 European Commission. Data protection day 2014: full speed on EU data protection reform [Internet]. Brussels: European Commission; [last updated 2014 May 13; cited 2014 Aug 1]. Available from: [http://europa.eu/rapid/press-release\\_MEMO-14-60\\_en.htm](http://europa.eu/rapid/press-release_MEMO-14-60_en.htm)
- 49 Hicks A, Henriksen JH, Koenig B. Involving the public in planning for the genomics revolution: an experiment in deliberative democracy. *Biotech Exam*. 2008;11(2):1–3.