



European Commission  
Directorate-General for Health and Consumers  
Unit D3 eHealth and Health Technology Assessment

# **The Use of Big Data in Public Health Policy and Research**

## **Background information document**

Brussels, 29 August 2014

# Index

1. Introduction .....	3
2. Background .....	3
2.1. The Big Data paradigm .....	3
2.2. Trends shaping the potential of Big Data.....	5
2.3. Big Data opportunities in public health and clinical research .....	6
2.4. Big Data challenges .....	8
2.4.1. Interpretation, Propensity, Correlations.....	8
2.4.2. Standards and Interoperability .....	9
2.4.3. Data Governance and Trust .....	9
2.4.4. Data Expertise and Infrastructure.....	10
3. Actions .....	10
3.1. Examples of Big Data activities in EU Member States .....	10
3.2. Examples of international activities using Big Data .....	11
3.3. Initiatives taken forward by the European Commission.....	13
4. References .....	17

## **1. Introduction**

The present background is focused on the *Use of Big Data in Public Health and Research* and is structured in three main chapters – 2. Background; 3. Actions; 4. Discussion and Conclusion.

In chapter 2, the Big Data paradigm and the trends shaping its potential will be identified. Subsequently, the Big Data opportunities in public health policy and research will be outlined in light of the logic of improvement of healthcare systems and research. The explanation of the challenges currently permeating the full potential of Big Data will close the opening chapter. Afterwards, some examples of Big Data activities in EU Member States and outside the EU as well as the initiatives taken forward by the European Union will be mentioned in chapter 3.

## **2. Background**

### **2.1. The Big Data paradigm**

A defining characteristic of today's data-rich society is the collection, storage, processing and analysis of immense amounts of data. This characteristic is cross-sectorial, ranging from the domain of machine learning and engineering, to economics and medicine [1].

One of the most commonly recognized applications of Big Data is social media data analysis, probably due to the extensive press coverage about social network impact on customer behaviour and experiences. It would be easy to conclude that Big Data means social media data, but this assumption would miss to capture both existing applications as well as the potential of the Big Data paradigm [2].

Big Data is generated from an increasing plurality of sources, including internet clicks, mobile transactions, user-generated content, and social media as well as purposefully generated content through sensor networks or business transactions such as sales queries and purchases. In addition, genomics, health care, engineering, operations management, the industrial internet, and finance all add to the Big Data pervasiveness [3]. But what does Big Data really mean?

Until now, there is no single **definition** of Big Data. Certainly, the absence of one sole definition signals the challenging landscape of its operation and meaning. Several definitions have been brought to the fore. According to McKinsey the term "Big Data" refers to *datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyse* [4]. Another popular definition of Big Data is the "3V" model proposed by Gartner: Big Data is *high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making* [5]. From a strategic and managerial point of view some argue that such definition needs further refinement. In this sense Big Data is also characterized according to a fourth dimension: Veracity, concerning the quality, authenticity, 'trustworthiness' of data [6] [7].

Another definition of a Big Data set is the one where "N = All" - where there is no need for sampling because the entire background population is available; and when "N = All" there is no issue of sampling bias because the sample includes everyone [8]. Furthermore, among practitioners, there is emergent discussion that "big" is no longer the defining parameter, but rather how "smart" the data is, focusing on the insights that the volume of data can reasonably provide [3].

Whether it is "big", "smart", "N = All", "3 or 4Vs", the use of large-scale data to predict human behaviour is gaining value in business and government policy practice, as well as in the scientific domain.

In health care, the complexity of Big Data analysis also arises from combining different types of information. Starting with the collection of individual data elements and moving to the fusion of multiple data sets, the results can reveal entirely new approaches to treating diseases [9].

For the purpose of this report, Big Data encompasses information from electronic health care records, social media, patient summaries, genomic and pharmaceutical data, test results, claims, telemedicine, mobile apps, home monitoring, clinical trials, sensors and information on wellbeing, behaviour and socioeconomic indicators.

Governments and public institutions are moving towards Open Data policies. Big Data and Open data are closely related but are not the same. Open data is accessible public data that people, companies, and organisations can use to launch new ventures, analyse patterns and

trends, make data-driven decisions and solve complex problems. All definitions of open data include two basic features: the data must be publicly available for anyone to use, and it must be licensed in a way that allows for its reuse. There is a general agreement that open data should be available free of charge or at minimal cost. Both Big Data and open data can transform business, government, research and society – and a combination of the two is especially potent. Big Data gives us unprecedented power to understand, analyse, and ultimately change the world we live in, while open data ensures that power will be shared. Ultimately, with the right framework conditions, the re-use of open and Big Data will be a motor for innovation, growth and transparency [10] [11].

## **2.2. Trends shaping the potential of Big Data**

Over the past few years, nearly all **major businesses** (e.g. Oracle, IBM, Microsoft, Google, Amazon and Facebook) have started their Big Data projects. Taking IBM as an example, since 2005, IBM has invested USD 16 billion on 30 purchases related to Big Data [12].

In **academia**, Big Data is also receiving great attention. In 2008, Nature published a Big Data special issue [13] and, in 2011, Science launched a special issue on the key technologies of “data processing” and the subjects surrounding the increasingly huge influx of research data [14]. In 2012, European Research Consortium for Informatics and Mathematics (ERCIM) News published a special issue on Big Data [15]. In the beginning of 2012, a report titled 'Big Data, Big Impact' announced that Big Data has become a new kind of economic assets, just like currency or gold [16]. In addition, the Journal of Biomedical and Health Informatics has announced a special issue on “Big Data for Health” in early 2015 with the scope of promoting Big Data as a source of innovation in healthcare and accelerating the translational pathways from the laboratory to the patient. The issue will address the latest technical development and practical applications of Big Data for health [17].

Big Data plays an important role in **medical and clinical research** having already contributed to clinically relevant studies and applied in disease research. For example, in gastroenterology [18], cancer [19] [20], dementia [21], disease prevention [22], epidemiology [23] [24], infectious disease surveillance [25], rare diseases [26] [27], and in the monitoring of kidney transplants [28]. The rising number of studies with large scope in the past few years

demonstrates the increasing desire of researchers to manipulate clinical Big Data and to translate discoveries into healthcare practice.

Furthermore, **researchers and funders** recognize the value of integrating clinical research networks. Connecting existing networks means clinical research can be conducted more effectively, ensuring that patients, providers, and scientists form true communities of research in an environment of shared operational knowledge and data. Major research institute centres and funding agencies have made large investments in this domain. E.g., the Li Ka Shing Centre for Health Information and Discovery of the University of Oxford was recently supported by a £90m initiative in Big Data and drug discovery [29] and the NIH Big Data to Knowledge (BD2K) initiative is enabling biomedical scientists to capitalize more fully on the Big Data being generated by the research communities [30].

### **2.3. Big Data opportunities in public health and clinical research**

The last two decades have seen an explosion in Big Data throughout the health-care value chain, as well as the advent of new platforms, tools, and methodologies in storing, structuring, and analysing Big Data. Important developments include the use of genomic data in drug discovery, the sharing of clinical-trial data, the use of electronic healthcare records (EHRs), and the increased availability of data from mHealth applications, patient registries, and social media [31]. The concept of data fusion is gaining further significance since on top of the collection of individual data elements arises the fusing together of multiple data sets [32].

The huge amount of novel data being generated is already making important contributions in epidemiology, more specifically in temporal **public health surveillance** [33]. In the era of satellite sensors, a diversity of epidemiologically relevant environmental information can be sourced globally at daily intervals. Big Data allows a closer temporal matching of disease outbreaks with covariates which may improve the accuracy of mapping models. The possibility of seasonally tailored geographic baselines certainly improves traditional temporal surveillance by facilitating early warning of epidemiologically relevant environmental changes [25].

The concept of the **quantified self**, in which individuals deploy sensors and monitoring devices to measure their own health and behaviour, has become a reality. If expanded and aggregated at population level, will lead to a data-driven approach of collective health and wellbeing measurement [22].

The generation of new knowledge about the effectiveness of treatments and the prediction of outcomes are of two fundamental applications of big healthcare data. In other words, Big Data can be applied in **prospective and retrospective research**.

**Prospective research** is outcome focused (e.g., development of a disease relative to other factors such as suspected risk or protection factor(s)) and has a low recall error as it involves longitudinal observations over time and results are collected at regular time intervals. Furthermore, Big Data differs from tradition decision support tools as it allows the collection and analysis real-time patient data [34]. New possibilities for the discovery of **innovative pharmaceuticals**, development of **more effective treatment protocols** and for the development of **personalised medicine** come into play.

Statistical tools and algorithms improve clinical trial design and patient recruitment to better match treatments to individual patients, reducing trial failures and speeding new treatments to the market. The analysis of clinical trials and patient records allows the identification of follow-up indications and the discovery of adverse effects before products reach the market, which ultimately will improve pharmacovigilance and patient safety [35].

The promise of personalized medicine guided by an understanding of each individual's genome has been fostered by increasingly powerful and economical methods to acquire clinically relevant information [36]. Big Data helps translate personalised medicine into clinical practice by offering the opportunity to use analytical capabilities that can integrate systems biology (e. g., genomics) with data from EHRs [34]. The integration of population science with individual genomic measurements will enable the practice of personalized medicine[37].

Big Data can also be applied **retrospectively** to evaluate the **effectiveness of treatments and protocols** and their **impact on quality of life**, helping reduce waste and inefficiency in clinical operations (determine more clinically relevant and cost-effective diagnosis and treatments), research and development[35]. In the clinical context, Big Data can also be used to determine causality, effect, or association between risk factors and the disease of interest.

From a payment perspective, Big Data can be used to ensure that providers have strong performance records and are reimbursed on the quality of patient outcomes rather than the quantity of care delivered[38].

Unquestionably, turning large amounts of data into actionable information allows the identification of needs, the provision of services, the prediction and prevention of disasters, especially for the benefit of populations.

## **2.4. Big Data challenges**

Big Data technologies can derive value from large datasets in ways that were previously impossible, but the technical capabilities of Big Data have reached a level of sophistication and pervasiveness that demands consideration of how to balance the opportunities afforded by Big Data against challenges that these technologies raise. In this respect, the following challenges were identified (the catalogue is by no means exhaustive):

### **2.4.1. Interpretation, Propensity, Correlations**

The opportunities to learn and generate value from Big Data systems will depend on the statistically valid use of the information. The size and heterogeneity of the data being collected is a major challenge, particularly since the majority of statistical approaches to interpretation were developed in an era when “sample sizes” were relatively small, and when data acquisition technologies and computing power were limited [39]. Nowadays, the high volume, velocity and variety of data collection methods available is likely to drive the data-driven society to a point in which sampling will not be necessary because the entire background population is available. By working with almost all the information about the phenomena there is a growing capacity to expand research questions.

In addition, Big Data encompasses a high level of messiness in the sense that the increase in the amount of information by orders of magnitude means giving up the preference for highly curated data for the sake of having a higher sample and effect size [40]. Although issues of sampling bias might decline because the sample includes everyone, what will be the cost of having more comprehensiveness? Are we likely to see patterns where none actually exist



simply because enormous quantities of data can offer connections that radiate in all directions (i.e., fear of apophenia)? [41] Or, on the contrary, the possibility to collect vastly more data than before and to process it less costly and quickly will allow the identification of the right combination of tweaks capable of bringing marginal changes that, when multiplied by a large number of instances, or allowed to work over a long time, will produce a significant effect? [42] Will society penalize infractions before they occur solely based on analytical predictions of future behaviours?

Structured data, such as tables of numbers, do not reveal everything that is known about a medication or biological process and much of what is known about living organisms exists in unstructured formats [9]. The analysis of unstructured, voluminous and disorganized data has brought significant discoveries [43] [44]. Nevertheless, in the absence of causality and strong theories, is society likely to fall blindly on correlations? [45]

### **2.4.2. Standards and Interoperability**

There are still standardization problems in the healthcare sector, as data is often fragmented, or generated in IT systems with incompatible formats [46]. Research, clinical activities, hospital services, education, and administrative services are siloed, and, in many organizations, each silo maintains its own separate organizational (and sometimes duplicated) data and information infrastructure. The lack of cross-border coordination and technology integration calls for standards to facilitate interoperability among the components of the Big Data value chain. Global standardization can contribute to addressing such challenges, thus which steps are required to fully realize it?

### **2.4.3. Data Governance and Trust**

As the amount of health related data and global digital information grows, so does the number of actors accessing and using this information. Assurances must be given so that personal data related to health will be used appropriately, in the context of the intended uses and according to the relevant laws. There is still scepticism with regards to “where the data goes to”, “by whom it is used” and “for what purpose” in the EU fragmented and overly complex legal environment. In what concerns to privacy, conditions under which data are

shared for research are being discussed under the proposed Data Protection Regulation [47] . How can we ensure Big Data is viewed as a national and Pan-European resource made responsibly available to those who can derive social value from it?

The discussion around the reliability of de-identification (i.e., storing and sharing the data without revealing the identity of the individuals involved) remains strong. On one hand, according to Cavoukian and Castro the findings of specific cases of re-identification cannot be overstated and de-identification is an effective mechanism for protecting personal information [48]. On the other hand, Narayanan and Felten criticize Cavoukian and Castro by arguing that de-identification still fails to resist inference of sensitive information and that attempts to quantify its feasibility promote a false sense of security [49]. Which techniques of data privacy and security techniques shall be promoted?

#### **2.4.4. Data Expertise and Infrastructure**

Big Data offers enormous possibilities for new insights, for understanding human systems at the systemic level, and for detecting interactions and nonlinearities in relations among variables. Nevertheless, traditional data analytics, insufficient infrastructure and funding opportunities, lack of trust in databases [50] and shortage of data experts [51] and related skills hinder the development of innovative data management solutions.

### **3. Actions**

#### **3.1.Examples of Big Data activities in EU Member States**

Regulators are beginning to use Big Data. **National Health Systems** have started to make patient-level data (e.g. prescriptions, diagnoses, procedures, physician visits, hospitalizations, lab tests) available through governments and payers. For instance, the **Danish** government provides researchers access to anonymous data (including health indicators and hospitalization data) on individual patients from the 1970s to the present day [31]. Moreover, Denmark has deployed a program using Big Data analytics to improve the quality and comprehensiveness of care for patients with chronic illness, namely diabetes and heart disease [52].

The Parliament of the **United Kingdom** recently issued a research briefing stating the importance of the conversion of health records into electronic form and the use of data (GP records of individual patients' illnesses and treatments, and data from all English NHS hospitals about all attendances, diagnoses and treatments) to grasp new opportunities for public health management and research [53].

**Sweden** is also in a strong position in Big Data analytics. In the healthcare sector, report systems have the capacity to merge data streams from lab results, ambulances and healthcare centres. Moreover, high quality applied research is enabling industrial actors to leverage on large data sets in collaboration with research institutes and academia. For example, frontline research is undertaken in health care analytics, which aims for providing efficient and effective decision support for health care and pharmaceutical research. This includes the development of techniques and tools to support decision making and discovery of drug effects by analysing structured and unstructured data [54].

In **Estonia** all public infrastructure is service oriented. One of the key elements of the Estonian national databases is the decentralization of data. There is no single data owner or controller, every government agency or business are able to choose the product that is right for them and digital services are added on the basis of need [55]. Data is stored where it is collected and is exchanged. The Estonian eHealth system was launched in 2008, ePrescriptions are used since 2010. Moreover, hospitals issue digital birth certificates and health insurance starts automatically [56].

In short, several countries already provide interesting examples of how data can be used to support innovation in health research and more productive health care systems. Countries that now plan how they will harness the value of health data will have the opportunity to become global leaders in health care innovation, with the resultant benefits of high performing health care systems and growth in health care innovation [57].

### **3.2.Examples of international activities using Big Data**

In January 2014, the **White House**, led by President Obama's Counsellor John Podesta, undertook a 90-day review of Big Data and privacy. The review was designed to outline what

was new about the technologies that define the Big Data landscape; uncover where and how Big Data affects public policy and the laws and norms governing privacy; ask how and whether Big Data creates new challenges for the principles of the Consumer Privacy Bill of Rights; and to set an agenda on how government can maximize the benefits and minimize the risks of Big Data. The 90-day review brought concrete recommendations [38] [58], namely:

- Policy attention should focus more on the actual uses of Big Data and less on its collection and analysis. Policies focused on the regulation of data collection, storage, retention, a priori limitations on applications, and analysis (absent identifiable actual uses of the data or products of analysis) are unlikely to yield effective strategies for improving privacy. Such policies would be unlikely to be scalable over time, or to be enforceable by other than severe and economically damaging measures.
- Policy concerning privacy protection should address the purpose (the “what”) rather than prescribing the mechanism (the “how”).
- Research (and funding for it) is needed in the technologies that help to protect privacy, in the social mechanisms that influence privacy preserving behaviour, and in the legal options that are robust to changes in technology and create appropriate balance among economic opportunity, national priorities, and privacy protection.
- Increased education and training opportunities concerning privacy protection, including career paths for professionals. Programs that provide education leading to privacy expertise are essential and need encouragement.
- Privacy Protections should be extended to non-U.S. citizens as privacy is a worldwide value that should be reflected in how the federal government handles personally identifiable information from non-U.S. citizens [59].

The White House underlined how the use of electronic health records (EHRs) vastly expanded the volume of health data available to clinicians, researchers and patients; the emerging practice of personalised medicine, as the ultimate application of Big Data in health; and the role of Big Data in ensuring professionals' stronger performance records and reimbursement on the quality of patient outcomes rather than on the quantity of care

delivered. Furthermore, the opportunities associated with the detection of symptoms by mobile devices and the risks of data governance and trust were mentioned [58].

In the domain of mHealth, the **World Health Organization** has launched the project “Be He@lthy Be mobile” in Senegal and under the mDiabetes initiative it supports countries to set up large-scale projects that use mobile technology, in particular text messaging and apps, to control, prevent and manage non-communicable diseases such as diabetes, cancer and heart disease [60]. mDiabetes is the first initiative to take advantage of the widespread mobile technology to reach millions of Senegalese people with health information and expand access to expertise and care. The project also includes a training module for health workers and will allow for remote consultations and monitoring of patients in rural areas. Launched in 2013, the initiative is also working on an mCessation for tobacco programme in Costa Rica, an mCervical cancer programme in Zambia and has plans to roll out mHypertension and mWellness programmes in other countries.

The **OECD Health Care Quality Indicators (HCQI)** project is responsible for a plan in 2013/2014 to develop tools to assist countries in balancing data privacy risks and risks from not developing and using health data. This plan includes developing a risk categorisation of different types and uses of data and the promising practices that countries can deploy to reduce risks and enable data use [57].

### **3.3. Initiatives taken forward by the European Commission**

Big Data technology and services are expected to grow worldwide at a compound annual growth rate of 40% [61]. Moreover, in the UK alone, the number of Big Data staff specialist working in large firms will increase by more than 240% over the next five years (ref). In the public sector, the progress in the IT environment (availability of broadband and Big Data tools, cloud services, HPC) will lead to cost reduction of operations, increase of efficiency and better and more personalised services for citizens[62].

As a response to the European Council’s conclusions of October 2013, the Commission adopted the Communication on the **data-driven economy**, which focused on the digital economy, innovation and services as drivers for growth and jobs and called for EU action to provide the right framework conditions for a single market for Big Data and cloud computing

[63]. The Communication describes the data-driven economy of the future and sets out operational conclusions to support and accelerate the transition towards it having the scope of stimulating research and innovation on data, creating more business opportunities and the availability of knowledge and capital across Europe will increase.

Moreover, the European Commission's strategy on the **data value chain** aims at extracting the maximum value from data to provide value for the economy and citizens, by building a data community across the Member States [64]. The key policy targets of the data value chain strategy are the following: increase the number of data-related jobs; increase the number of data-related start-ups and fast-growing SMEs; increase the revenue generated based on data in the Member States; improved use of data for decision-making processes in the private sector and the public sector; and increase citizens' use of data for informed behavioural decisions.

According to a Staff Working Document jointly prepared by DG SANCO and DG RTD, the development of personalised medicine through the use of –omics technologies offers new opportunities for the treatment of patients [65]. Namely, better targeted treatments, less medical errors and a reduction of adverse reactions to medicinal products. The European Commission has continued to monitor the developments of personalised medicine and, in a recent press release proposed the creation of Big Data public-private partnerships to fund personalised medicine [66]. Furthermore, the ongoing Horizon 2020 Big Data call for proposals may result in funding further projects in the field of public health and clinical research.

The political guidelines for the next European Commission announced by President Jean-Claude Juncker are focused on the creation of a connected single market in a vibrant knowledge-based society [67]. Ambitious legislative steps towards a connected digital single market shall be taken to safeguard the fundamental right of data protection within the European Union and externally, namely with the United States.

An important condition for the access to patient-related information is the protection of personal data. The data protection rules as transposed into national laws in Europe do not yet establish full harmonized conditions for health data processing. Such fragmentation creates costs and disincentives for science and business. The Commission is currently in the process

of negotiating the general EU legal framework on the protection of personal data. The proposed **General Data Protection Regulation** aims at eliminating fragmentation and providing consistency and coherence for the whole Union [47]. The proposed Regulation has specific provisions on data processing for health purposes and on historical, statistical and scientific research purposes. The main policy objectives for the Commission are to modernise the EU legal system for the protection of personal data, in particular, to meet the challenges resulting from globalisation and the use of new technologies; the strengthening of individuals' rights; the reduction of administrative formalities to ensure a free flow of personal data within the EU and beyond; and the improvement of the clarity and coherence of the EU rules for personal data protection.

The Commission launched a public consultation with the **Green Paper on mobile health** focused on the barriers and issues related to the use of mHealth [68]. The objective of the Green Paper, announced in the eHealth Action Plan 2012-2020, was to launch a broad stakeholder consultation on existing barriers and issues related to mHealth deployment and help identify the right way forward to unlock the mHealth potential. Regarding Big Data, the consultation aimed at collecting evidence on the following question: 'What measures are needed to fully realise the potential of mHealth generated "Big Data" in the EU whilst complying with legal and ethical requirements?' The consultation was open from 10 April to 10 July 2014. The results of the consultation and the feedback report will be issued in the fourth quarter of 2014. On the basis of the responses, the Commission may take steps to support the implementation of mHealth in the EU, to be announced in 2015.

DG SANCO is committed to capturing the potential of Big Data in public health policy and research and to produce policy recommendations to Member States according to the logic improvement of healthcare systems and in light of the **Directive 2011/24/EU on Patients' Rights in Cross-Border Healthcare** [69]. The activities of the eHealth and the Health Technology Assessment Networks shall benefit from the recommendations on Big Data.

Article 14 of the Directive on patients' rights institutes the legal policy cooperation on eHealth at EU level and the objectives of the **eHealth Network**. With the growing possibilities to use the huge amounts of data stored in health data-bases and health records, questions regarding how to make effective use of this data and how public authorities will protect the information of patients, have emerged. Several Member States already see the

public discussion rising, and there is a push for new legislation. The eHealth Network Members exchange information and experiences regarding the activities of their public authorities. The eHealth Network Multi-Annual Work Plan 2015-2018 defines concrete actions to be taken in the domains of interoperability and standardisation; exchange of knowledge; assessment of implementation; and global cooperation and positioning cross the European Union [70]. The adoption of guidelines on the electronic exchange of patient summaries, in November 2013 is one example of the deployment activities of the eHealth Network [71].

The Commission plans to launch a follow-up of the **PATients REGistries iNiTiative Joint Action** (PARENT) that shall also benefit from the developments in Big Data. The overall objective of PARENT is to support Member States in developing comparable and coherent patient registries in fields of identified importance (e.g. chronic diseases, rare diseases, medical technology) with the aim to rationalize the development and governance of patient registries, thus enabling analyses of secondary data for public health and research purposes [72]. As a follow-up on the 2015 health programme, the focus will be on medical devices and pharmacovigilance needs for post market launch of evidence on data collection and generation and will evaluate the long term and relative effectiveness of real-world data based on patient registries and **Electronic Healthcare Records** (EHRs). The advent of EHRs has also created new opportunities to generate and disseminate new knowledge through the secondary use of clinical data [73].



## 4. References

1. Wang, W. and E. Krishnan, *Big data and clinicians: a review on the state of the science*. Journal Medical Informatics, 2014. **2**(1): p. e1.
2. IBM, *Analytics: The real-world use of big data How innovative enterprises extract value from uncertain data*, in *Executive Report*. 2013, IBM Institute for Business Value: United States of America.
3. George, G., M.R. Haas, and A. Pentland, *Big Data and Management*. Academy of Management Journal, 2014. **57**(2): p. 321-326.
4. Institute, M.G., *Big data: The next frontier for innovation, competition, and productivity in Report*. 2011, McKinsey Global Institute.
5. Beyer MA, D.L., *The Importance of 'Big data': A Definition*, in *Gartner IT Glossary*. 2012.
6. McAfee, A. and E. Brynjolfsson, *Big data: the management revolution*. Harvard Business Review, 2012. **90**(10): p. 60-6, 68, 128.
7. Ferrando-Llopis, R., D. Lopez-Berzosa, and C. Mulligan, *Advancing value creation and value capture in data-intensive contexts*. IEEE 2013: p. 5-9.
8. Harford, T. (2014) *Big data: are we making a big mistake?* . Financial Times Magazine.
9. May, M., *Life Science Technologies: Big biological impacts from big data*. Science, 2014. **344**(6189): p. 1298-1300.
10. Gurin, J. (2014) *Big data and open data: what's what and why does it matter?* The Guardian.
11. Commission, E., *Open data - An engine for innovation, growth and transparent governance* E. Commission, Editor. 2011, European Commission: Brussels.
12. Chen, M., S. Mao, and Y. Liu, *Big Data: A Survey*. Mobile Networks and Applications, 2014. **19**(2): p. 171-209.
13. *Community cleverness required*. Nature, 2008. **455**(7209): p. 1-1.
14. *Special Issue: Dealing with Data*. Science, 2011. **331**(6018).
15. News, E.R.C.f.I.a.M., *Special theme: Big Data*, in *European Research Consortium for Informatics and Mathematics News* M.K.C.a.C.T.I.-C. Stefan Manegold, Editor. 2012.
16. Forum, T.W.E., *Big Data, Big Impact: New Possibilities for International Development*. 2012: Geneva.
17. Informatics, J.o.B.a.H. *Call-for-Papers: Special issue on "Big Data for Health" in early 2015*. Journal of Biomedical and Health Informatics 2014; Available from: <http://jbhi.embs.org/wp-content/uploads/sites/3/2014/04/jbhi-big-data.pdf> accessed 15 July 2014
18. Genta, R.M. and A. Sonnenberg, *Big data in gastroenterology research*. Nature Reviews Gastroenterology & Hepatology, 2014. **11**(6): p. 386-90.
19. Meyer, A.M., et al., *Big data for population-based cancer research: the Integrated Cancer Information and Surveillance System*. N C Med J, 2014. **75**(4): p. 265-9.
20. Savage, N., *Bioinformatics: Big data versus the big C*. Nature, 2014. **509**(7502): p. S66-7.
21. *Addressing global dementia*. Lancet, 2014. **383**(9936): p. 2185.
22. Barrett, M.A., et al., *Big data and disease prevention: From quantified self to quantified communities*. Big Data, 2013. **1**(3): p. 168-175.
23. Khoury, M.J., et al., *Transforming epidemiology for 21st century medicine and public health*. Cancer Epidemiol Biomarkers Prev, 2013. **22**(4): p. 508-16.
24. Kao, R.R., et al., *Supersize me: how whole-genome sequencing and big data are transforming epidemiology*. Trends in Microbiology, Cell press, 2014. **22**(5): p. 282-91.
25. Hay, S.I., et al., *Big data opportunities for global infectious disease surveillance*. PLoS Med, 2013. **10**(4): p. e1001413.
26. Berman, J., *Rare Diseases and Orphan Drugs Keys to Understanding and Treating the Common Diseases*. 10 June 2014: Academic Press.

27. Ekins, S., et al., *Bigger data, collaborative tools and the future of predictive drug discovery*. J Comput Aided Mol Des, 2014.
28. Krivov, S.V., et al., *Optimal reaction coordinate as a biomarker for the dynamics of recovery from kidney transplant*. PLoS Comput Biol, 2014. **10**(6): p. e1003685.
29. Oxford, U.o. *Prime Minister joins Sir Ka-shing Li for launch of £90m initiative in big data and drug discovery at Oxford University*. 3 May 2014; Available from: [http://www.ox.ac.uk/media/news\\_releases\\_for\\_journalists/130305.html](http://www.ox.ac.uk/media/news_releases_for_journalists/130305.html).
30. BD2K, N.B.D.t.K. *Mission Statement*. 2012; Available from: [http://bd2k.nih.gov/about\\_bd2k.html#sthash.p9yaYnX7.dpbs](http://bd2k.nih.gov/about_bd2k.html#sthash.p9yaYnX7.dpbs) accessed 20 June 2014.
31. Szlezak, N., et al., *The role of big data and advanced analytics in drug discovery, development, and commercialization*. Nature, 2014. **95**(5): p. 492-5.
32. Bernstein, S.W., *Q&A:The US big data report and fully utilising big data within healthcare*. eHealth Law & Policy, 2014. **1**(4).
33. Salathe, M., et al., *Digital epidemiology*. PLoS Comput Biol, 2012. **8**(7): p. e1002616.
34. Murdoch, T.B. and A.S. Detsky, *The inevitable application of big data to health care*. JAMA, 2013. **309**(13): p. 1351-1352.
35. Raghupathi, W. and V. Raghupathi, *Big data analytics in healthcare: promise and potential*. Health Information Science and Systems, 2014. **2**(1): p. 3.
36. Pulley, J.M., et al., *Operational implementation of prospective genotyping for personalized medicine: the design of the Vanderbilt PREDICT project*. Clinical Pharmacology & Therapeutics, 2012. **92**(1): p. 87-95.
37. Altman, R., *Personal genomic measurements: the opportunity for information integration*. Clinical Pharmacology & Therapeutics, 2013. **93**(1): p. 21-23.
38. Podesta, J., et al., *Big data: seizing opportunities, preserving values*. Executive Office of the President. Report. Retrieved on May, 2014. **1**: p. 2013.
39. OECD, *Strengthening Health Information Infrastructure for Health Care Quality Governance*. 2013: OECD Publishing. 180.
40. Mayer-Schönberger, V. and K. Cukier, *Big data: A revolution that will transform how we live, work, and think*. 2013: Houghton Mifflin Harcourt.
41. Boyd, D. and K. Crawford, *Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon*. Information, Communication & Society, 2012. **15**(5): p. 662-679.
42. Schumpeter (2014) *Little things that mean a lot*. The Economist.
43. Baker, E.W., *Relational Model Bases: A Technical Approach to Real-time Business Intelligence and Decision Making*. Communications of the Association for Information Systems, 2013. **33**(1): p. 23.
44. White, R.W., et al., *Web-scale pharmacovigilance: listening to signals from the crowd*. Journal of the American Medical Informatics Association, 2013: p. amiajnl-2012-001482.
45. Lazer, D.M., et al., *The parable of Google Flu: Traps in big data analysis*. Science, 2014. **343** (6176): p. 1203-1205.
46. Roney, K. (2012) *If Interoperability is the Future of Healthcare, What's the Delay?* . Becker's Hospital Review.
47. Commission, E., *Proposal for a Regulation on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) in COM(2012) 11 final* 2012: Brussels.
48. Cavoukian, A. and D. Castro, *Big Data and Innovation, Setting the Record Straight: De-identification Does Work*. Discussion Paper - PHIPA, 2014.
49. Narayanan, A. and E.W. Felten, *No silver bullet: De-identification still doesn't work*. 2014.
50. Agaku, I.T., et al., *Concern about security and privacy, and perceived control over collection and use of health information are related to withholding of health information from*

- healthcare providers*. Journal of the American Medical Informatics Association, 2014. **21**(2): p. 374-378.
51. Davenport, T.H. and D. Patil, *Data Scientist*. Harvard Business Review, 2012. **90**: p. 70-76.
  52. IBM (2013) *Southern Denmark Tackles Heart Disease with IBM Big Data Analytics Program*. Press Release.
  53. Border, P. (July 2014) *Big Data and Public Health* UK Parliament.
  54. Network, T.S.B.D.A., *Big Data Analytics A Research and Innovation Agenda for Sweden*. 2013, The Swedish Big Data Analytics Network Sweden.
  55. e-Estonia.com (2001) *X-Road*.
  56. Economist, T. (2014) *Estonia takes the plunge A national identity scheme goes global*. The Economist.
  57. OECD, *Data-Driven Health Care Innovation, Management and Policy*, DELSA/HEA(2013)13, Editor. 2013, OECD.
  58. PCAST, P.s.C.o.A.o.S.a.T., *Big Data and Privacy: a Technological Perspective*, W. House, Editor. 2014: Washington.
  59. House, W. (2014) *FACT SHEET: Big Data and Privacy Working Group Review*.
  60. WHO (2014) *Mobile phones help people with diabetes to manage fasting and feasting during Ramadan*. Features 2014.
  61. Vesset, D., et al., *Worldwide Big Data Technology and Services 2012–2015 Forecast*. IDC Report, 2012. **233485**.
  62. UK, e.-s., *Big Data Analytics: An assessment of demand for labour and skills, 2012-2017 e.-s.* UK, Editor. 2013, SAS UK: London.
  63. Commission, E., *Towards a thriving data-driven economy 2014*, COM(2014) 442 final Brussels.
  64. Commission, E., *A European strategy on the data value chain* D. CONNECT, Editor. 2013, European Commission.
  65. Commission, E., *Use of '-omics' technologies in the development of personalised medicine*, in *Commission Staff Working Document*, S. final, Editor. 2013, European Commission: Brussels.
  66. Commission, E., *Commission urges governments to embrace potential of Big Data* in *Press Release*, P. Release, Editor. 2014: Brussels.
  67. Juncker, J.-C. (2014) *A New Start for Europe: My Agenda for Jobs, Growth, Fairness and Democratic Change*.
  68. Commission, E., *Green Paper on mobile Health ("mHealth")*, C. final, Editor. 2014, European Commission: Brussels.
  69. Union, E.P.a.C.o.t.E., *DIRECTIVE 2011/24/EU on the application of patients' rights in cross-border healthcare*. 9 March 2011.
  70. Commission, E., *Multi-Annual Work Plan 2015-2018*, e.N.S.-g.o.P.f.M.-A.W.P. 2015-2018, Editor. 2014.
  71. Commission, E., *Guidelines on Minimim/Non-Exhaustive Patient Summary Dataser for Electronic Exchange in Accordance with the Cross-Border Directive 2011/21/EU*. 2013: Brussels.
  72. PARENT. <http://www.eunethta.eu/parent>. Available from: <http://www.eunethta.eu/parent>.
  73. Danciu, I., et al., *Secondary use of clinical data: the Vanderbilt approach*. Journal of biomedical informatics, 2014.
  74. Commission, E. ( July 2014) *Commission encourages re-use of public sector data* Press Release.