Health Systems & Reform

ISSN: 2328-8604 (Print) 2328-8620 (Online) Journal homepage: http://www.tandfonline.com/loi/khsr20

# Big Data and Health Care: Challenges and Opportunities for Coordinated Policy Development in the EU

Sebastian Salas-Vega, Adria Haimann & Elias Mossialos

**To cite this article:** Sebastian Salas-Vega, Adria Haimann & Elias Mossialos (2015) Big Data and Health Care: Challenges and Opportunities for Coordinated Policy Development in the EU, Health Systems & Reform, 1:4, 285-300, DOI: 10.1080/23288604.2015.1091538

**To link to this article:** https://doi.org/10.1080/23288604.2015.1091538

Accepted author version posted online: 16 Sep 2015.
Published online: 16 Sep 2015.

Submit your article to this journal

Article views: 8314

View Crossmark data

Citing articles: 5 View citing articles

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=khsr20

*Health Systems & Reform*, 1(4):285–300, 2015
Copyright © Taylor and Francis Group, LLC
ISSN: 2328-8604 print / 2328-8620 online
DOI: 10.1080/23288604.2015.1091538

Taylor & Francis
Taylor & Francis Group

## Research Article

# Big Data and Health Care: Challenges and Opportunities for Coordinated Policy Development in the EU

Sebastian Salas-Vega*, Adria Haimann and Elias Mossialos
London School of Economics and Political Science; London, UK

**CONTENTS**



**Abstract**—As global policy makers prioritize big data policy, it is important to try to outline expected outcomes vis-à-vis health sector objectives. We identify initiatives aimed at promoting the use of big data in European Union (EU) health care, highlight expected challenges, and use these to evaluate EU big data policy developments to the extent that they are able to advance health sector priorities. A comprehensive approach is used to capture and examine peer-reviewed and gray literature publications on the use of big data in global health systems. This approach involved electronic database and specialist website searching, as well as complementary use of search engines and qualitative inputs from key EU policy stakeholders. Ongoing health data initiatives revolve around data center development, confidentiality and security, e-health and m-health, and genomics and bioinformatics. The literature acknowledges several main challenges to the successful integration of big data in health care, classified as either ethical (confidentiality and data security, access to information) or technical (data reliability, interoperability, data management and governance). EU data policy has started to address these issues, though additional work remains. A larger outstanding challenge is the lack of a comprehensive health and research policy strategy for big data that targets sectoral objectives. It remains unclear how big data integration will affect the quality and performance of health care in the EU. The promises of big data are being eroded by a failure to develop a coherent approach to adequately address conceptual, ethical, and technical challenges pertaining to its use within EU health systems.

## INTRODUCTION

"Big data" has become a popular theme in European Union (EU) policymaking. The European Commission (EC) recently inaugurated its "Digital Agenda for Europe" (DAE) as one of the seven pillars of the Europe 2020 initiative[1] to capitalize on the data revolution and foster innovation and economic growth throughout the Union. Calls to expedite the

Keywords: big data, Europe, health policy, qualitative research, reform

Received 19 March 2015; revised 1 September 2015; accepted 3 September 2015.

*Correspondence to: Sebastian Salas-Vega; Email: s.salas-vega@lse.ac.uk

creation of a single market for big data have since emerged from recent European Council Summits (October 2013), and the EC has responded by adopting legal, technical, and logistical aims for big data reforms in the EU.[2-4] Tangible outcomes from the EC's pivot to data policy reform include the publication of guidelines on open standard licenses, data sets, and reuse charging; as well as revisions to the Public Sector Information directive, making public data "open by default"[5] and growing the ecosystem for big data.[6,7] Perhaps the most visible outcome of big data policy reform was the signing of a contractual public–private partnership (cPPP) between the EC and industry (Big Data Value Association) on 13 October 2014, which has pledged €2.5 billion to "strengthen the data sector in Europe."[8]

Efforts by European policy makers to grow data availability, processing, and use in the EU appear to be driven by socioeconomic aspirations. Union publications repeatedly highlight expectations of data-driven social and economic gains for Europe,[2-5,8-10] yet provide limited detail on how they are to materialize.

The likelihood of data policy reforms engendering social and economic gains can arguably be estimated by referring to legislative goals and agendas. In health, key EU policy objectives include (**Table 1**) strengthening of health system effectiveness, accessibility, resilience, quality, and performance,[11] as well as the promotion of health research.[12] The extent to which data policy reforms advance these objectives in health provides a measure of the marginal benefit from ongoing political and financial investments in EU data.

Importantly, however, major conceptual dilemmas in successfully bridging big data policy and EU health care remain: first, there is no framework for measuring data-driven progress across EU member states toward common health policy and research objectives. Second, data policy developments have not been informed by conceptual issues that underlie sectoral objectives, such as how to best measure health system quality and performance. Third, health data collection processes across countries are not systematized, and comprehensive structures capable of incentivizing common progress remain undefined. Finally, the extent to which obstacles to integrated big data–health policymaking are compounded by poor coordination within EU political institutions remains unclear.

As a consequence, emerging EU initiatives on big data[5,13,14] have not been brought under a coherent conceptual framework that captures, coordinates, and advances health policy and research objectives. Recent communications on the push for data-driven economies have even acknowledged that sectoral priorities for data research and innovation remain unidentified,[2] even as data policy implementation continues. Set beside health system improvement initiatives that presuppose availability of data, as well as communication across performance assessment and publication agendas, it is surprising that robust EU policies cutting across data and health remain lacking.

EU health policy stands at a critical juncture: greater action must be taken to help carefully weave the fabric of data-laden health care across the EU. Failure to do so risks (1) development of crude data programs that are incapable of adequately addressing the needs of health systems and (2)

| EU Health Sector Objectives | Scope of Interest to EU Policy Makers |
|---|---|
| • Effectiveness | A key component of health care quality and health system performance, effectiveness is defined as health systems' ability to produce positive health outcomes across the population. Data on comparative effectiveness of health systems can be informed by EU-wide indicators. |
| • Accessibility | A fundamental tenet of European health care that is characterized by ease of access to medical treatment. EU states are obligated to administer socially inclusive health systems. |
| • Resilience | Health systems must be resilient to both fiscal and nonfiscal challenges; that is, they must be able to quickly adapt to unexpected challenges in supply and demand. The EC is tasked with supporting member states in this work through analysis, forecasts, and reform recommendations. |
| • Research and innovation | EU research and data analysis is prioritized to help tackle health afflictions and societal challenges and promote learning. Some research is of such scale and complexity that it requires EU level coordination to improve efficiency in use of resources to address common challenges, such as improving health. |

**TABLE 1.** Key EU Health Sector Objectives and Scope of Interest to EU Policy Makers. Source: Adapted from the European Commission.[11,12]

continued fragmentation in EU data strategies,[3] driven by a lack of common vision for reform.

To maximize the benefit of further data policy reforms to the health sector, policy must be able to exploit synergies in the merger between health and data. A critical examination of the interaction between both fields is therefore needed. To address this issue, this article takes a comprehensive approach to capture ongoing EU initiatives and challenges regarding data use in health. These are evaluated alongside EU policy developments to provide a realistic assessment of the scope for data-driven improvements to health.

We find that EU policy makers have begun to tackle technical and ethical issues regarding data use, though with noticeably less attention to focused applications in health. Policy developments have also yet to address many of the conceptual challenges that underlie attainment of health sector ambitions. In light of these gaps, we conclude that the EU should reevaluate its strategy for data policy development to help meet sectoral objectives and maximize the value drawn from its investments.

## DEFINING BIG DATA

### Existing Definitions

This article does not set out to review definitions of big data, as these are already available elsewhere.[15-17] Suffice it to say that no single definition of big data is universally accepted,[18] though certain definitions do stand out. Big data, for instance, is most commonly defined by the 3 Vs (volume, velocity, and variety),[19-23] as well as other variants: there are the "4 Vs," which captures data volume, velocity, variety, and veracity[24,25] and the "5 Vs," which also considers value.[26,27] Big data can otherwise be defined as a large collection of complex data sets,[28-32] characterized by the existence of structured and unstructured variable sets.[33,34]

### EU Definitions

Nonsystematic use of big data definitions also extends to the EU policy arena. The EC describes big data as data that is "difficult to process with current data management tools and methods."[2] The same source, however, also refers to big data in terms of the 3 Vs: "Large amounts of different types of data produced with high velocity from a high number of various types of sources."[2] Defining big data nonsystematically undermines the establishment of normative standards to apply in regulation and may also give rise to internal inconsistencies—one EC definition identifies big data through its development, the second through its processing.

In the absence of an a priori reason to do otherwise, we define big data through a merged EC definition: a large amount of different types of data produced with high velocity from various types of sources and which must be processed through novel approaches to bypass processing limitations extending from current management tools and methods.

## METHODS

A comprehensive approach—consisting of electronic database and specialist website searching, as well as complementary use of search engines and qualitative inputs from key policy stakeholders—was used to capture the peer-reviewed and non-peer-reviewed literature commenting on big data use in health care.[35] From this, we critically summarize active data initiatives and identify overarching principles and challenges associated with their merger to health. These are set beside a discussion of ongoing big data policy reforms in the EU to shed light on the extent to which they can realistically promote European health objectives. Appropriate strategies for coordinated data–health development are then identified and discussed.

## DATA SOURCES AND SEARCH STRATEGY

A detailed search strategy was used in July 2014 to massively expand peer-reviewed publications indexed through Medline via PubMed, Scopus, Google Scholar, and EconLit. The search strategy included logic string combinations of relevant text words, keywords, and medical subject headings, developed through internal consensus among researchers and further refined with input from research librarians (**Table 2**). A hierarchical procedure for database searching was used to independently capture the literature discussing big data in health care (first search level) and that discussing policies pertaining to big data in health care (second search level). Only relevant hits published in English since 2005 were included.

Two assessors independently conducted title and abstract screening of all captured articles to identify those that focused on big data in US or EU health care. Where there was disagreement on inclusion eligibility ($n = 12$), consensus was reached through discussion. Papers discussing technical big data computing techniques, detailed methods for big data statistical analysis, and cost analyses of big data integration were beyond the scope of this article and were thus excluded.

| String | Terms |
|---|---|
| String 1: big data (search level: 1st + 2nd) | PM: (big data [tiab]) OR (big data [tw])<br>GS/EL: ("big data")<br>S: ("big data" [title/abs/keyw]) |
| AND | |
| String 2: health care (search level: 1st + 2nd) | PM: (healthcare [tiab]) OR (health care) OR (health) OR (health care system) OR (delivery of health care [mh])<br>GS: (health [title]) OR (healthcare [title]) OR ("health care" [title]) OR ("health care system" [title])<br>S: (healthcare [title/abs/keyw]) OR (health [title/abs/keyw]) OR ("health care system" [title/abs/keyw]) |
| AND | |
| String 3: policy (search level: 2nd) | PM: (polic* [tiab]) OR (legislation [tiab]) OR (legislation [mh]) OR (legislat* [tiab])<br>GS: (policy) OR (policies) OR (legislation) OR (legislative)<br>S: (polic*[title/abs/keyw]) OR (legislation [title/abs/keyw]) OR (legislat* [title/abs/keyw]) |

**TABLE 2.** Literature Review Search Terms (Search Levels). PM, PubMed; GS, Google Scholar; EL, EconLit, S, Scopus

All papers meeting selection criteria were included in the review ($n$ = 164).

The search strategy was also extended to capture the gray literature. Full and modified search strategies were applied within relevant organization websites (e.g., EC, Eurostat, European Parliament) and search engines (Google). Literature searches were also complemented by publication recommendations and qualitative inputs from two content experts and key policy stakeholders at Eurostat (EC). By capturing peer-reviewed and non-peer-reviewed publications, this article incorporates viewpoints on the use of big data in health care from academia, government, and industry.

This approach is used to combine different perspectives on the technical, ethical, and conceptual challenges that pertain to the use of big data within and across health systems. We preface this synthesis with a discussion of ongoing big data initiatives in the EU to help assess the adequacy of recent big data developments vis-à-vis health sector needs and characteristics.

## RESULTS

This section is divided into two components: first, ongoing big data initiatives in health across the EU are reviewed, at both national and European levels. To the extent that EU policymaking exists to manage and coordinate national policy efforts, prominent challenges associated with additional development of big health data arrangements are then identified and discussed within the context of EU policy. We conclude that despite the proliferation of policy targeting big data use in the EU, significant challenges remain to its successful application in EU health care.

## ONGOING INITIATIVES

Big data in health is a broad theme that can cover a wide array of topics. A discussion of ongoing initiatives across the EU is therefore developed on the basis of a four-tiered thematic framework encompassing topics debated in influential US and EU big data publications: data centers,[24,36] confidentiality and data security,[24,25,37] e-health and m-health,[24,25,37] and genomics and bioinformatics.[24]

### Data Centers

Countries are vigorously attempting to develop data centers—databases that can store vast amounts of diverse health information—for the purpose of clinical practice and research,[7,18,38-52] public health surveillance,[53-57] medical training and learning,[58-62] and pharmaceutical development and marketing.[63-71]

The literature suggests that research in oncology, cardiology, neurology, mental, and population health is benefiting from the growth of big data,[38,39,50,51] with the discussion on cancer predominating.[40-48] The quantity of generated cancer data is rapidly increasing in the EU and the United States, particularly through tumor genome sequencing, computed tomography and magnetic resonance imaging, test results, and medical history.[42] In some instances, these data are being compiled into cancer registries to give insights into timing of diagnosis, as well as long-term patient outcomes and treatment effects.[40] In Italy, IBM has partnered with an Italian cancer institute to leverage available data with the aim of improving cancer care.[45] In the UK, Oxford University has established its Big Data Institute and the Chan Soon-Shiong Oxford Center for Molecular Medicine to collect and analyze large anonymized medical datasets and promote data-driven

personalization in cancer medicine.[49,51] Also in the UK, Public Health England and the National Cancer Intelligence Network have moved to build the world's largest database of cancer patients to coordinate and develop analysis and intelligence and improve oncology prevention, treatment, and outcomes.[48] At a European level, the recently launched Innovative Medicines Initiative 2 has issued calls to fund the development of knowledge repositories to enhance personalization of patient care, and future topics seek to promote the development of value-based health data systems in the EU.[52] While clinical registries and longitudinal patient health surveys are increasingly being consolidated, the extent to which these data sets are being utilized to effectively address health needs remains less clear.

Big data research centers are also being used for public health surveillance. Researchers and health officials are using aggregated individual data to monitor global disease trends in real time.[53-56] Visual analytics are also being used to help choose appropriate intervention policies on the basis of foodborne illness trends.[57]

Big data centers are also being integrated into medical training and educational initiatives.[58-62] Outside of the EU, health professional education is using longitudinal data related to curriculum structure to determine appropriate competencies and course trajectories.[59]

Additionally, pharmaceutical data pools are being leveraged for pharmacovigilance,[65-67,71] monitoring of developmental life cycles,[63,64] and pharmaceutical marketing and commercialization.[69] In Europe, big data has been used to help with pharmaceutical development through the Innovative Medicines Initiative, an EU-funded public–private program intending to pool data and accelerate medicinal development.[70]

There are indications that the pharmaceutical industry stands to gain from greater integration of health data,[72] yet opposing claims also exist. There is the view that increased data access will facilitate drug safety and efficacy assessment,[73] yet others have suggested that added data rigors—for example, the establishment of health information protection systems—may disadvantage industry by increasing legal, reputational, and financial risks.[74]

## Confidentiality and Data Security

Strategies for confidentiality and data security are key focus areas in the big data literature. In terms of policy, a number of EC communications[2,75] and directives[5,76] have begun to shape a common legal framework targeting patient confidentiality concerns by establishing effective data protection and network and information security rules. More recently, the EC adopted a proposal for a general EU framework for data protection in January 2012 with a view to modernize and harmonize data protection rules.[77] The European Data Protection Directive complements this by prohibiting data processing that may infringe fundamental freedoms or privacy, except if data subjects give their explicit consent or if circumstances permit derogation.[76] The EC has also recommended that information and communication technologies integrate the principle of privacy-by-design and default and make use of privacy-enhancing technologies.[77,78] The extent to which these nonbinding guidelines have been adopted nevertheless remains unclear.

At an operational level, US health care industries are currently trying to tackle privacy challenges through the creation of secure data clouds[22,79-82] that make use of privacy mechanisms such as obfuscation[80,83,84] to safeguard health confidentiality. Privacy-preserving record linkage and analytics—in which algorithms run on encrypted data—are also being developed to decouple personal and sensitive information (e.g., cancer status), helping to maintain patient anonymity and thus enabling its protected use in research.[31,85,86] Because digital security is a globally relevant issue with replicable, adaptable, and interdependent solutions,[87,88] the EU likely benefits from global developments in big data security.

Big data in health often involves the aggregation of patient-level information for secondary use. To address public concerns regarding confidentiality and misuse of collected personal data[89] that underpin legal and ethical frameworks for its use, informational campaigns may be used to raise awareness of data privacy rights and protection mechanisms. The EC has stated that it intends to use educational campaigns to inform the public on ways to reduce confidentiality and data security risks,[2] though operational details remain lacking. National authorities appear to be taking a more proactive role in this regard: the English National Health Service, for instance, recently led a public campaign to inform citizens about care.data, a program making wider use of deidentified health information for research purposes,[90] though indications suggest that it failed.[91] It is unclear how the EU and member states have moved to effectively coordinate informational campaigns.

## e-Health and m-Health

Several articles emphasize that big data is expected to improve disease management by better informing individualized diagnosis and treatment.[29,92-95] One example of a personalized medicine initiative is a Danish theoretical service model called Co-production of Health, designed to unite health care and self-care to provide "health added value" by

computing through personalized models that are context-aware.[96] This model for health care targeting, however, remains unimplemented and unvalidated. Ongoing efforts to create pan-European biobanks—for example, the Central Research Infrastructure for Molecular Pathology and the Organisation for Economic Co-operation and Development's Global Biological Resources Centers Network—also offer an opportunity to individualize care by helping to reveal the disease relevance of genes.[94] For its part, the EU has moved to fund academic and industry initiatives in personalized medicine, such as PerMed and EuroBioForum, that promote data sharing.[77] Despite these initiatives, however, the full potential of big data in personalized medicine is unlikely to be met without parallel advances in regulatory, reimbursement, and privacy legislation.[97,98]

Emerging mobile and computer-based health applications have helped patients share personal treatment experiences and promote physical and mental health.[99-102] These applications are often combined with sensor systems that collect vast amounts of information and serve a variety of purposes, ranging from elderly assistance to informing overweight patients about increased health risks.[32,103-105] From this perspective, complex algorithms, artificial intelligence, and machine learning are needed for dynamic, secondary analysis of rapidly growing health data sets that increasingly link patient-generated data,[106-108] yet these issues are not reflected in EU policy. Technologies supporting real-time data collection and processing may be particularly useful in enabling the EU to effectively and quickly adapt to changing health environments.[11] Greater legislative clarity is required to effectively coordinate data-generating applications in health.[77]

Electronic health records (EHRs) similarly provide an abundance of data with potential value to clinical medicine.[33,47,109-113] Many EU countries, including The Netherlands, Denmark, and the UK, are introducing EHR systems that update individual health history following medical consultations or treatment.[114] These countries, have been advised by Eurostat (EC) to take the lead in investigating uses of such systems for statistical and big data analytical purposes.[114] Though this coincides with the EC's decision to support "lighthouse" data initiatives across the EU, it may prove challenging for the bloc to successfully integrate national EHR data to the benefit of health policy objectives regarding cross-country comparisons[115] without international management and coordination.

### Genomics and Bioinformatics

Genomics and bioinformatics is another key topic cross-linking big data and health. Two main uses of genomics include the sequencing of malignant tumors and genomes.[116] The amount of captured genetic data is rapidly growing as a result of next-generation technologies that use high-throughput DNA sequencing[117-119] to boost genetic profiling capacity.

Sequencing and translational bioinformatics[24,120-123] represent big data applications that require massive amounts of storage and analytical power for data processing. To accommodate, infrastructure and big data tools, including cloud computing and storage techniques, are being tailored for use, particularly in the genetic and genomic sciences.[116,120,124-130] An example of this in the EU is the Helix Nebula Project. This public–private initiative across information technology providers, the European Organization for Nuclear Research, the European Molecular Biology Lab, and the European Space Agency, uses cloud services to perform on-demand, large-scale genomic analysis.[131] In the UK, public–private partnerships are also investing in big data health research centers that focus on studying the early stages of disease using genomic and chemical screens.[51]

## CHALLENGES SURROUNDING USE OF BIG DATA IN HEALTH CARE

Several key challenges are frequently said to present an obstacle to big data use in health care: confidentiality and data security,[24,89] access to information,[132,133] data reliability,[34,134] interoperability,[135,136] and management and governance.[137,138] This section discusses these challenges within the context of EU big data policy and initiatives.

### Confidentiality and Data Security

Patients fear that misappropriation of their health information—particularly genetic data[34,139,140]—may adversely affect personal circumstances, including insurance coverage and employment.[25] Unfortunately, data access and confidentiality risks are directly correlated.[141] In excluding scientific and medical data from general principles making public data open-by-default, the EC's Open Data policy appears to reflect these concerns.[5] The organization has instead opted for complementary legislation to address unique confidentiality challenges in health[142] and cross-border care delivery.[143] Given that public approval is a chief regulator of the political will for reform, many of the remaining data policy challenges identified in this article may flow from this central point.

The EC has planned to further address patient confidentiality concerns through amendments to existing data protection directives,[76,89,144] following EU constitutional revisions that strengthen personal data protection rights (Treaty of Lisbon). These legislative changes are to unify EU initiatives on

confidentiality and data security and to provide a more flexible legal framework that can rapidly adapt to changing technologies. However, data protection reforms have arguably been few and limited to enhancing transparency[89] and confidentiality in lawful data processing. In the context of sensitive health data, it is unclear whether this promotes broader ambitions for data-driven health sector development, innovation, and private sector involvement.

### Access to Information

A competing challenge is access to information. Individual perceptions of powerlessness in data control are currently at odds with organizational beliefs of data ownership.[145] Concerns regarding data access and use legitimize the question: are society's best interests in mind as data access pathways are negotiated? For the consumer, a primary concern is third party access and data control.[139,140] Companies, too, are interested in internal collection and use of information[63] but also worry about disclosure of intellectual property.[146]

The financial and nonfinancial interests of health care providers may also be challenged by information sharing. These organizations may consequently be less inclined to disclose information regarding performance[132] or may actively work to game data return through exception reporting[147,148] or cherry-picking of patients,[149] giving rise to concerns of data validity.

Across the EU, several countries have made an effort to provide more information to patients on health-related issues. The European Collaboration for Health Optimization project—a Spanish initiative to collect health data and analyze variations in European medical practice and health outcomes—and the English care.data program serve as two examples.[150,151]

To expand data access, the EC has also adopted an open-by-default principle to public sector information, freely making it available for commercial and noncommercial reuse.[5] Though open data is not necessarily "big," it can be used to help accelerate the maturation of a big data ecosystem.[6,7] The amended Public Sector Information directive nevertheless excludes scientific and health-related information, diminishing its applicability to health data. Elsewhere, EU government agencies, academia, industry, nongovernmental organizations, and international organizations have moved to enhance clinical data sharing in an effort to enhance trial transparency and analytical reproducibility.[152-156]

### Data Reliability

Data reliability is another often-cited challenge to implementation and use of big data systems in health.[34,113,134]

At an operational level, manually fed electronic health data may be prone to error and bias from human entry. Yet, regularized systems can also introduce systematic bias into data collection and analysis.[28,34,133,134] For example, underfunded organizations lacking adequate technological infrastructures to document and share information may only be able to capture data from a subset of the population[34] or may in fact capture incorrect information if collection or processing algorithms are flawed.[113] As others have pointed out, if an EHR lacks information about a medical event, it is not necessarily because the event did not occur.[135] Blind acceptance of big data should therefore be cautioned against: there is a need for measured use of big data and careful interpretation of results, as well as investment in big data system development and audit.

EU member states are actively trying to improve data reliability. The UK Hospital Episode Statistics (HES), for instance, is regarded as one of the world's most comprehensive health data sets, processing over 125 million patient records per year for all admissions (1990–), outpatient appointments (2003–), and accident and emergency attendances (2008–) occurring within English public hospitals. Despite its tremendous value to research and clinical practice, HES has historically been criticized for unreliable secondary diagnoses.[157] In response, UK health authorities refocused efforts to improve clinical coding in 2001[157] and established an assurance program for hospital data in 2007.[158] Recent studies have reported improvements in HES data quality,[159] demonstrating the value of regulatory oversight in ensuring health data reliability. Despite EU health policy and research objectives that require a healthy source of data,[115] data quality initiatives remain lacking at an EU level.

### Interoperability

Data interoperability is another major challenge to further development of medical data systems.[132-135,160] Interoperability is crucial for recording health information, developing common interfaces, agreeing on common data sets, and defining quality standards.[114] Interoperability necessitates development of data platforms in an international, comparable context and thus requires common principles. Kenny Simmen, Vice President of Janssen's Infection Diseases, Research and Early Development, alluded to this point by explaining that different countries have different regulations, making it difficult for coordination and collaboration between researchers.[161] Interoperability of EU data sets is further complicated by varying clinical standards and languages.[134]

The EC's Digital Agenda for Europe highlights an opportunity to deliver sustainable economic and social benefits from a digital single market that is based on interoperable applications.[4] The EU recently sponsored the epSOS project to outline how member states can integrate e-health architectures and established the e-Health Network as the main strategic and governance body in the EU to work toward interoperability of cross-border e-health services.[77,143] However, participation in these networks remains voluntary,[143] and only this year will the EC—with the endorsement of the e-Health Network—propose an e-Health Interoperability Framework to help establish legal, organizational, semantic, and technical specifications for interoperable cross-border e-Health services.[77] Notably, the e-Health Interoperability Framework will include a non-exhaustive list of data to be collected through patient summaries and shared as part of cross-border data sharing initiatives to promote continuity of EU care.[77] This comes alongside a push by the EC to examine member states' laws on electronic health records in order to assess the legal aspects of interoperability.[77] In the absence of additional progress, it nevertheless remains unclear how these initiatives will enhance big health data interoperability across the Union and effectively promote health policy and research objectives.

### Management and Governance

There is little information in the published literature discussing governance and management of health data at the EU level. Big data leaders within the EC have indicated to us that DG Health and Food Safety (DG Sante), DG Communications Networks, Content and Technology (DG Connect), and DG Research and Innovation are the main EU governance bodies capable of informing big data debates in health care. Yet, there is little evidence of coordinated efforts to promote health sector objectives through the use of big data. Indeed, although European authorities are responsible for assembling comparable health-related data across member states and developing mechanisms for comparative analysis,[115] DG Sante and Eurostat have only recently begun to outline strategies to improve comparative health reporting.[162]

It also remains unclear how responsibilities regarding health data systems are split across European agencies. For example, although Eurostat and DG Sante are responsible for expanding and improving the European Health Survey System,[162] it is unclear how other relevant bodies (e.g., DG Connect) support this endeavor. From a user's perspective, improved clarity from the EU on health data governance and management is needed[138] and may facilitate access and effective data use.

## DISCUSSION

Our review of initiatives and challenges regarding big data in EU health care highlights four key lessons for big data policy development across the EU. We present these here alongside an evidence-based discussion of potential strategies for remediation. Though these are framed around EU health data and policy developments, they also provide globally relevant insights on how to develop capacity and coherent policy in digital health resources to promote domestic and international health policy initiatives.

## A BALANCED FRAMEWORK ENCOMPASSING DATA CONFIDENTIALITY AND USE

As EU policy makers begin to call for greater integration of health data sets,[163,164] it is important to bear in mind that as in any security domain, the weakest link can break the chain. Relegating policy responsibilities regarding health confidentiality and data security to individual member states may create a weak and non-uniform cross-border data privacy architecture. EU governing councils therefore have an important role to play in establishing the framework for common progress in confidentiality and data security.

Despite ongoing efforts by the EC to update data protection legislation to accommodate rapidly evolving technologies,[78,89] a more comprehensive and coherent policy on the fundamental right to personal data protection is needed.[144,165] This should balance confidentiality, data access and security in health, while also defining standards for data ownership and control, reuse, cross-border flow, storage, and processing.[77] Existing legal frameworks define personal data, establish principles for its lawful processing, and also strengthen individual access rights.[144] However, they may be at times vague, overly restrictive, and internally inconsistent: for instance, they may call for the removal of regulatory red tape, but at the same time prohibit the processing of genetic or health data without providing exclusion for qualified third-party use that nevertheless safeguards personal data protection.[144] Big data systems in health must be strategically designed to ensure patient confidentiality, while granting timely access to qualified users in academia and industry, to fully achieve health policy and research objectives.[141] The EC should take this into consideration as it finalizes its review of the EU legal framework for personal data protection, which aims to strengthen individuals' rights and facilitate commercial data use.[78]

As data policy reform is debated, it may be useful to consider stakeholder interests regarding data access. Health care professionals, researchers, and industry all have valid interests

in comprehensive access to big health data, such as improvement in quality of care, clinical trial development, and pharmacovigilance. On the other hand, health data access demands from formal (e.g., insurers) and informal (e.g. hackers) entities may be unacceptable. The EC should take this into consideration, especially as it fields nongovernmental partnerships to expedite data growth, processing, and use in the EU.[10]

## INTEGRATION OF DATA SYSTEMS

Centralized EU governance provides the bloc with an opportunity to be at the vanguard of multinational health data integration. Although several initiatives have started to address e-Health interoperability,[77] major outcomes are only expected this year and their impact on data program remains unknown. Many related issues—for example, researcher access to interoperable e-Health data and promotion of health policy objectives—also remain unresolved.

There is hope that big data integration will optimize preventative care by helping to address risk factors[92,99] and improve measurement of health system performance.[166] Cross-country comparisons of health information is in line with EU competencies, interests, and health policy and research objectives.[115] However, comparisons should be based on conceptual linkages between health and data use—for example, that leverage valid quality indicators—to ensure progress toward these objectives.

A particularly valuable opportunity for big data integration within the EU health context may be in the sharing of clinical evidence.[77] By leveraging across established EU health registries, well-designed and highly powered secondary studies may help attenuate traditional sources of bias,[167] better inform clinical applications (e.g., comparative effectiveness research), and thus promote member state and stakeholder interests in health technology assessment.[77]

Nevertheless, data quality and interoperability must first be promoted to ensure meaningful comparison. International data sets often reflect slightly different definitions of health and health care and may be developed using varying algorithms, making linkage difficult and of questionable quality. Existing differences in clinical coding and quality measuring practices,[168] for instance, can bias cross-country comparisons of hospital performance. It also remains difficult to produce high-quality, error-free linkages across health data sets even within highly developed health systems.[169]

With this as a backdrop, there is no EU-level body dedicated to data monitoring and integration efforts in health, even though cooperative health data collection, analysis, monitoring, and dissemination activities fall within the legal remit of European public health interests and strategies.[115] The Reform Treaty of 2007 in fact reaffirms the EC's role in comparative health policy, and both the EC and European Parliament have called for greater exchange of information by assembling comparable health data across member states and developing mechanisms for comparative analysis, including through multilevel health indicators.[115,170] Although a set of European core health indicators exist on health status, health determinants, and care across EU member countries, they remain less than perfectly comparable, data are not always available, and it is unclear how they correspond to recent data policy reforms.[171]

Finally, industry bears legitimate interest in promoting data quality improvements and closer integration within and across national boundaries.[172] European policy makers should take industry as a partner in ongoing big data reform initiatives, particularly as privately managed health data sets—often existing beyond the reach of public directives[5,13]—gain prominence.[173]

## ENSURING QUALITY OF DATA COLLECTION, ANALYSIS, AND OVERSIGHT

The EU is unlikely to capitalize on big data in health until there is high-quality data collection and processing systems and well-defined governance providing oversight.

Sitting within Europe's political and legislative hub, EU policy makers should give greater clarity to big data governance in health care, particularly as it applies to cross-border data use. The statistical service of the EU, Eurostat (EC), has already created a Big Data Task Force to refine use of big data for European official statistics. This task force, however, does not focus on data use in health but rather on its application to all EU statistics. Given the unique challenges associated with health data,[174] EU policy makers should consider creating expert teams to oversee EU health data quality initiatives.

Notably, the EU is currently witnessing the arrival of affordable personal data-generating devices[80,104] that may complement traditional streams of health data and enhance provider knowledge.[175] Broader policy developments regarding the extension of property rights to personal data[61] may, however, ultimately define collection and use of information available through this medium.

## UNTANGLING HYPE FROM REALITY

There is a tremendous amount of hype surrounding big data in health care. Some anticipate that big data will help enhance efficiency, quality, and equity in health care delivery,[25] whereas others expect big data to produce

tremendous cost savings.[82,176,177] The EC frequently adopts this rationale to justify European investments in data.

We do not dispute that big data has the *potential* to improve health and health care and lead to efficiency savings by better informing sector processes. However, to realize these expectations, conceptual, ethical, and technical concerns must first be overcome. Ongoing failures to tailor data policy to these issues points to a European strategy for big data use in health care that is not grounded in a coherent vision for policy development. These are likely to have created an unconnected patchwork of effects in health care across member states,[178] and they draw to question whether widely acknowledged expectations for big data will be met, at least insofar as EU health policy and research objectives are concerned.

## CONCLUSIONS

European policy makers have started to develop policy affecting the European data market. Policy developments reflect recognition within the EC of the need for a common big data strategy and infrastructure. In this article, we examined the extent of big data policy coordination with broader EU health policy and research objectives. Policy developments have started to address technical (e.g., interoperability) and ethical (e.g., legal frameworks regarding confidentiality and data security) challenges to the use of big data in health care, two major barriers to e-Health development identified in EU action plans.[77] However, EU policy makers have yet to tailor data policy to accommodate conceptual challenges to health sector development—for example, quality and performance improvement—that fall within European legal competencies and responsibilities in health.

Nevertheless, EC discussions on big data policy are still in their infancy, as confirmed by communications with key EU policy stakeholders. Additional progress in the merger between big data policy and sectoral objectives may therefore be expected in the near future as the EC embarks on this new field of policy. At this time, however, it remains unclear how big data developments will advance health sector objectives, casting doubt on optimistic predictions of the return on big data investments in the EU.

## DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

No potential conflicts of interest were disclosed.

## ACKNOWLEDGMENTS

We would like to thank two anonymous sources within Eurostat (EC) for their insights and feedback on big data policy developments within the EU.

## FUNDING

This research was funded by LSE Health.

## REFERENCES

[1] European Commission. Communication from the Commission: a strategy for smart, sustainable and inclusive growth. 2010. COM/2010/2020. Available at http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52010DC2020&from=EN

[2] European Commission. Communication from the Commission: towards a thriving data-driven economy. 2014. COM/2014/442. Available at http://ec.europa.eu/newsroom/dae/document.cfm?action=display&doc_id=6210

[3] European Commission. Communication from the Commission: open data: an engine for innovation, growth and transparent governance. 2011. COM/2011/882. Available at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0882:FIN:EN:PDF

[4] European Commission. Communication from the Commission: a digital agenda for Europe. 2010. COM/2010/245. Available at http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52010DC0245R(01)&from=EN

[5] European Commission. Amendment of the directive on the re-use of public sector information. 2013. Directive 2013/37/EU. Available at http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32013L0037&from=EN

[6] Buchholtz S, Bukowski M, Sniegocki A. Big & open data in Europe: a growth engine or a missed opportunity? 2014. Available at http://www.bigopendata.eu (accessed 24 January 2015)

[7] Oberlander J. Big data and data science in Scotland: an SSAC discussion document. 2014. Available at http://www.scottishscience.org.uk/sites/default/files/article-attachments/SSAC%20Big%20Data%20and%20Data%20Science%20in%20Scotland%20Paper%20March%202014_0.docx (accessed 24 January 2015)

[8] European Commission. Press release: European commission and data industry launch €2.5 billion partnership to master big data. 2014. IP/14/1129. Available at http://europa.eu/rapid/press-release_IP-14-1129_en.pdf

[9] European Commission. Memo: making the most of the data-driven economy. 2014. MEMO/14/455. Available at http://europa.eu/rapid/press-release_MEMO-14-455_en.pdf

[10] European Commission. Press release: commission urges governments to embrace potential of big data. 2014. IP/14/769. Available at http://europa.eu/rapid/press-release_IP-14-769_en.pdf

[11] European Commission. Communication from the Commission: on effective, accessible and resilient health systems.

2014. COM/2014/215. Available at http://ec.europa.eu/health/healthcare/docs/com2014_215_final_en.pdf

[12] European Commission. Press release: research & innovation: commission calls for partnerships to tackle societal challenges. 2011. IP/11/1059. Available at: http://europa.eu/rapid/press-release_IP-11-1059_en.pdf

[13] European Commission. Directive on the re-use of public sector information. 2003. Directive 2003/98/EC. Available at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:345:0090:0096:en:PDF

[14] Cabinet Office. Explanatory memorandum to the re-use of public sector information regulations 2005. 2005; No. 1515. Available at http://www.legislation.gov.uk/uksi/2005/1515/memorandum/contents

[15] MIT Technology Review. The big data conundrum: how to define it? 2013. Available at http://www.technologyreview.com/view/519851/the-big-data-conundrum-how-to-define-it/ (accessed 24 January 2015)

[16] De Mauro A, Greco M, Grimaldi M. What is big data? A consensual definition and a review of key research topics. In: 4th International Conference on Integrated Information: 2014 Sept 5-8; Madrid, Spain. AIP Conf Proc 2015; 1644:97-104. Available at: http://scitation.aip.org/content/aip/proceeding/aipcp/10.1063/1.4907823 (accessed 24 January 2015)

[17] Gandomi A, Haider M. Beyond the hype: big data concepts, methods, and analytics. Int J Inf Manage 2015; 35(2): 137-144.

[18] Halamka JD. Early experiences with big data at an academic medical center. Health Aff 2014; 33(7): 1132-1138.

[19] Howie L, Hirsch B, Locklear T, Abernethy AP. Assessing the value of patient-generated data to comparative effectiveness research. Health Aff 2014; 33(7): 1220-1228.

[20] Jee K, Kim GH. Potentiality of big data in the medical sector: focus on how to reshape the healthcare system. Healthc Inform Res 2013; 19(2): 79-85.

[21] Margolis R, Derr L, Dunn M, Huerta M, Larkin J, Sheehan J, Guyer M, Green ED. The National Institutes of Health's Big Data to Knowledge (BD2K) initiative: capitalizing on biomedical big data. J Am Med Inform Assoc 2014; 21(6):957-958.

[22] Roski J, Bo-Linn GW, Andrews TA. Creating value in health care through big data: opportunities and policy implications. Health Aff 2014; 33:1115-1122.

[23] High-Level Group for the Modernisation of Statistical Production and Services. What does "big data" mean for official statistics. 2013. Available at http://www1.unece.org/stat/platform/download/attachments/77170614/Big%20Data%20HLG%20Final%20Published%20Version.docx?version=1&modific (accessed 24 January 2015)

[24] Bellazzi R. Big data and biomedical informatics: a challenging opportunity. Yearb Med Inform 2014; 9(1):8-13.

[25] Feldman B, Martin EM, Skotnes T. Big data in healthcare: hype and hope. 2012. Available at http://www.scribd.com/doc/107250349/Big-Data-in-Healthcare-Hype-and-Hope#scribd (accessed 17 August 2014)

[26] Hanlon A. Big data and the 5Vs. 2014. Available at http://hiveintelligence.com/big-data-5vs/ (accessed 24 January 2015)

[27] Morley-Fletcher E. Big data healthcare: an overview of the challenges in data intensive healthcare. 2013. Available at http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=3499 (accessed 17 August 2014)

[28] Fihn SD, Francis J, Clancy C, Nielson C, Nelson K, Rumsfeld J, Cullen T, Bates J, Graham GL. Insights from advanced analytics at the veterans health administration. Health Aff 2014; 33(7): 1203-1211.

[29] Fleurence RL, Beal AC, Sheridan SE, Johnson LB, Selby JV. Patient-powered research networks aim to improve patient care and health research. Health Aff 2014; 33(7): 1212-1219.

[30] Nambiar R, Bhardwaj R, Sethi A, Vargheese R. A look at challenges and opportunities of big data analytics in healthcare. In: 2013 IEEE International Conference on Big Data: 2013 Oct 6–9; Santa Clara, CA, USA.

[31] Patil HK, Seshadri R. Big data security and privacy issues in healthcare. In: 2014 IEEE International Congress on Big Data; 2014 June 27–July 2; Anchorage, AK, USA.

[32] Yoon JP. Three-tiered data mining for big data patterns of wireless sensor networks in medical and healthcare domains. In: ICIW: The Eighth International Conference on Internet and Web Applications and Services; 2013 June 23-28; Rome, Italy.

[33] Poulymenopoulou M, Malamateniou F. Machine learning for knowledge extraction from PHR big data. In: Integrating information technology and management for quality of care, Mantas J, Househ MS, Hasman A, eds. Amsterdam, The Netherlands: IOS Press; 2014.

[34] Fallik D. For big data, big questions remain. Health Aff 2014; 33(7): 1111-1114.

[35] EPPI-Centre. EPPI-Centre methods for conducting systematic reviews. 2010. Available at http://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=hQBu8y4uVwI%3D&tabidD88 (accessed 24 January 2015)

[36] Groves P, Kayyali B, Knott D, Van Kuiken S. The "big data" revolution in healthcare: accelerating value and innovation. Center for US Health System Reform Business Technology Office; 2013. Available at http://www.pharmatalents.es/assets/files/Big_Data_Revolution.pdf

[37] Bollier D. The promise and peril of big data. 2010. Available at http://citpsite.s3.amazonaws.com/events/big-data/Aspen-Big_Data.pdf (accessed 17 August 2014)

[38] Lupse O, Crisan-Vida M, Stoicu-Tivadar L. Supporting diagnosis and treatment in medical care based on big data processing. In: Cross-border challenges in informatics with a focus on disease surveillance and utilising big data, Stoicu-Tivadar L, De Lusignan S, Orel A, Engelbrecht R, Surján G, eds. Amsterdam, The Netherlands: IOS Press, 2014.

[39] Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research. Gen Hosp Psychiatry 2010; 35(4): 332-338.

[40] Kanza G, Babic A. Data mining in cancer registries: a case for design studies. In: IFMBE Proceedings, Roa Romero LM, ed. New York, NY: Springer; 2014.
[41] Kuriyan J, Cobb N. Forecasts of cancer & chronic patients: big data metrics of population health. arXiv 2013. Available at http://arxiv.org/abs/1307.3434
[42] Savage N. Bioinformatics: big data versus the big C. Nature 2014; 509(7502): S66-S67.
[43] Schilsky RL, Michels DL, Kearbey AH, Yu PP, Hudis CA. Building a rapid learning health care system for oncology: the regulatory framework of CancerLinQ. J Clin Oncol 2014; 32(22): 2373-2379.
[44] Shaikh AR, Butte AJ, Schully SD, Dalton WS, Khoury MJ, Hesse BW. Collaborative biomedicine in the age of big data: the case of cancer. J Med Internet Res 2014; 16(4): e101.
[45] EHealthNews. Analytics and big data improve cancer treatment at Italian institute. 2013. Available at http://www.ehealthnews.eu/ibm/3519-analytics-and-big-data-improve-cancer-treatment-at-italian-institute (accessed 24 January 2015)
[46] McNeil C. CancerLinQ prototype demonstrates use of big data in clinical care. ASCO Post 2013; 4(7). Available at http://www.ascopost.com/issues/may-1,-2013/cancerlinq-prototype-demonstrates-use-of-big-data-in-clinical-care.aspx
[47] Yu P, Artz D, Warner J. Electronic Health Records (EHRs): Supporting ASCO's vision of cancer care. In: ASCO Educational Book. Vol. 34, Dizon D, ed. Alexandria: American Society of Clinical Oncology, 2014; 225–231.
[48] Gallagher J. Public Health England to launch largest cancer database [internet]. BBC News. 2013. Available at http://www.bbc.co.uk/news/health-22870352 (accessed 24 January 2015)
[49] Hodson J. Chan Soon-Shiong Institute for Molecular Medicine and the University of Oxford establish partnership to create the first center for genomic and proteomic medicine in the United Kingdom to benefit NHS cancer patients. Reuters. 2014. Available at http://uk.reuters.com/article/2014/10/23/cssiomm-idUKnBw235784a+100+BSW20141023 (accessed 24 January 2015)
[50] Twachtman G. Big data destined for the bedside within 5 years. Clinical Neurology News Digital Network. 2014. Available at http://www.acssurgerynews.com/news/news/single-article/big-data-destined-for-the-bedside-within-5-935years/93a7ce09df79331c4a0da5c34e02305b.html (accessed 17 August 2014)
[51] Gibney E. Oxford big data centre to get 30 million. Times Higher Education. 2013. Available at http://www.timeshighereducation.co.uk/news/oxford-big-data-centre-to-get-30-million/2003661.article (accessed 24 January 2015)
[52] European Commission. Innovative Medicines Initiative 2. 2014. Available at http://www.imi.europa.eu/content/imi-2 (accessed 24 January 2015)
[53] Hay SI, George DB, Moyes CL, Brownstein JS. Big data opportunities for global infectious disease surveillance. PLoS Med 2013; 10(4): e1001413.
[54] Nash DB. Harnessing the power of big data in healthcare. Am Health Drug Benefits 2014; 7(2): 69-70.
[55] Ambord C, Favre F, Faeh D, Chiolero A. Public health surveillance with big data: assessing diabetes trends using medico-administrative data. Eur J Public Health 2014; 24(2) Suppl: 353-354.
[56] Parliamentary Office of Science and Technology. Big data and public health. 2014. Available at http://researchbriefings.files.parliament.uk/documents/POST-PN-474/POST-PN-474.pdf (accessed 24 January 2015)
[57] Ola O, Sedig K. The challenge of big data in public health: an opportunity for visual analytics. Online J Public Health Inform 2014; 5(3): 1–21.
[58] Krumholz HM. Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. Health Aff 2014; 33(7): 1163-1170.
[59] Ellaway RH, Pusic MV, Galbraith RM, Cameron T. Developing the role of big data and analytics in health professional education. Med Teach 2014; 36(3): 216-222.
[60] Etheredge LM. Rapid learning: a breakthrough agenda. Health Aff 2014; 33(7): 1155-1162.
[61] Harper E. Can big data transform electronic health records into learning health systems? In: Nursing Informatics 2014, Saranto K, Weaver CA, Chang P, eds. Amsterdam, The Netherlands: IOS Press; 2014.
[62] Pentland A, Reid TG, Heibeck T. Big data and health: revolutionizing medicine and public health. 2013. Available at http://kit.mit.edu/sites/default/files/documents/WISH_BigData_Report.pdf (accessed 24 January 2015)
[63] Schultz T. Turning healthcare challenges into big data opportunities: a use-case review across the pharmaceutical development lifecycle. Bull Assoc Inf Sci Technol 2013; 39 (5): 34-40.
[64] Seebode C. BIG DATA infrastructures for pharmaceutical research. In: 2013 IEEE International Conference, 2013 Oct 6-9; Silicon Valley, CA, USA. Available at http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&arnumber=6691759
[65] Abbott R. Big data and pharmacovigilance: using health information exchanges to revolutionize drug safety. Iowa Law Rev 2013; 99(1): 225-292.
[66] Gombocz E. Changing the model in pharma and healthcare—can we afford to wait any longer? in: Data Integration in the life sciences, Baker CJO, Butler G, Jurisica I, eds. New York: Springer; 2013.
[67] Lio X, Chen H. AZDrugMiner: an information extraction system for mining patient-reported adverse drug events in online patient forums. In: Smart health, Zeng D, Yang CC, Tseng VS, Xing C, Chen H, Wang FY, Zheng X, eds. New York: Springer; 2013.
[68] Szlezák N, Evers M, Wang J, Pérez L. The role of big data and advanced analytics in drug discovery, development, and commercialization. Clin Pharmacol Ther 2014; 95(5): 492-495.
[69] Wolin B. Big data, pharmaceutical marketing and healthcare. J Digit Soc Media Mark 2014; 2(1): 35-39.
[70] European Commission. Innovative Medicines Initiative. 2014. Available at http://www.imi.europa.eu/content/mission (accessed 17 August 2014)
[71] Simpson SE, Madigan D, Zorych I, Schuemie MJ, Ryan PB, Suchard MA. Multiple self-controlled case series for large-scale longitudinal observational databases. Biometrics 2013; 69(4): 893-902.

[72] Darrow JJ, Avorn J, Kesselheim AS. New FDA breakthrough-drug category: implications for patients. N Engl J Med 2014; 370(13): 1252-1258.

[73] Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014; 2(3): 1-10.

[74] Emam K, Waldo A, Wright C. Webinar: Fear and loathing of data monetization [internet]. Privacy Analytics. 2014. Available at http://www.privacyanalytics.ca/wp-content/uploads/2014/04/Fear-and-Loathing-Data-Monetization-FINAL.pdf (accessed 24 January 2015)

[75] European Commission. Communication from the Commission: safeguarding privacy in a connected world: a European data protection framework for the 21st century. 2012. COM/2012/09. Available at http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52012DC0009&from=en

[76] European Commission. Directive on the protection of individuals with regard to the processing of personal data and on the free movement of such data. 1995. Directive 95/46/EC. Available at http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046&from=en

[77] European Commission. Communication from the Commission: eHealth Action Plan 2012–2020. 2012. COM/2012/736. Available at http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52012DC0736&from=EN

[78] European Commission. Protection of personal data. 2015. Available at http://ec.europa.eu/justice/data-protection/ (accessed 24 January 2015)

[79] Fabian B, Ermakova T, Junghanns P. Collaborative and secure sharing of healthcare data in multi-clouds. Inf Syst 2014; 48(2015): 132-150.

[80] Hsieh J-C, Li AH, Yang CC. Mobile, cloud, and big data computing: contributions, challenges, and new directions in telecardiology. Int J Environ Res Public Health 2013; 10 (11): 6131-6153.

[81] Sahoo SS, Jayapandian C, Garg G, Kaffashi F, Chung S, Bozorgi A, Chen CH, Loparo K, Lhatoo SD, Zhang GQ. Heart beats in the cloud: distributed analysis of electrophysiological "big data" using cloud computing for epilepsy clinical research. J Am Med Inform Assoc 2014; 21(2): 263-271.

[82] Youssef AE. A framework for secure healthcare systems based on big data analytics in mobile cloud computing environments. Int J Ambient Syst Appl 2014; 2(2): 1-11.

[83] Mowbray M, Pearson S, Shen Y. Enhancing privacy in cloud computing via policy-based obfuscation. J Supercomput 2012; 61(2): 267-291.

[84] Pearson S, Shen Y, Mowbray M. A privacy manager for cloud computing. In: Cloud computing, Jaatun MG, Zhao G, Rong C, eds. New York: Springer; 2009.

[85] Gehrke J. Quo vadis, data privacy? Ann N Y Acad Sci 2012; 1260(1): 45-54.

[86] Kum HC, Krishnamurthy A, Machanavajjhala A, Reiter MK, Ahalt S. Privacy preserving interactive record linkage (PPIRL). J Am Med Inform Assoc 2014; 21(2): 212-220.

[87] Hare F. Borders in cyberspace: can sovereignty adapt to the challenges of cyber security. In: The virtual battlefield: perspectives on cyber warfare, Czosseck C, Geers K, eds. The virtual battlefield: perspectives on cyber warfare. Amsterdam, The Netherlands: IOS Press; 2009.

[88] Jansen van Vuuren J, Leenen L, Zaaiman J. Using an ontology as a model for the implementation of the national cybersecurity policy framework for South Africa. In: 9th International Conference on Cyber Warfare & Security: 2014 Mar 24-25; West Lafayette. Available at http://www.researchgate.net/publication/265361934_Using_an_Ontology_as_a_Model_for_the_Implementation_of_the_National_Cybersecurity_Policy_Framework_for_South_Africa

[89] Rubinstein IS. Big data: the end of privacy or a new beginning? Int Data Priv Law 2013; 3(2): 74-87.

[90] Ganesh J. Big Data may be invasive but it will keep us in rude health. Financial Times. 2014 Feb 21. Available at http://www.ft.com/cms/s/0/62a5aaaa-9a55-11e3-8232-00144feab7de.html#axzz3qKZ9TXJK (accessed 17 August 2014)

[91] Triggle N. Care.data: How did it go so wrong? 2014. Available at http://www.bbc.co.uk/news/health-26259101 (accessed 24 January 2015)

[92] Chawla NV, Davis DA. Bringing big data to personalized healthcare: a patient-centered framework. J Gen Intern Med 2013; 28(3): S660-S665.

[93] Dilsizian SE, Siegel EL. Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. Curr Cardiol Rep 2014; 16(1): 1-8.

[94] Asslaber M, Zatloukal K. Biobanks: transnational, European and global networks. Brief Funct Genomics Proteomics 2007; 6(3): 193-201.

[95] Medtech Media. Leveraging big data and analytics in healthcare and life sciences: enabling personalized medicine for high-quality care, better outcomes. 2012. Available at http://www.intel.co.uk/content/dam/www/public/us/en/documents/white-papers/healthcare-leveraging-big-data-paper.pdf (accessed 17 August 2014)

[96] Boye N. Co-production of health enabled by next generation personal health systems. In: pHealth 2012, Blobel B, Pharow P, Sousa F, eds. Amsterdam, The Netherlands: IOS Press; 2012.

[97] Ginsburg GS, Willard HF. Genomic and personalized medicine: foundations and applications. Transl Res 2009; 154(6): 277-287.

[98] European Commission. Staff working document: use of "-omics" technologies in the development of personalised medicine. 2013. SWD/2013/436. Available at: http://ec.europa.eu/health/files/latest_news/2013-10_personalised_medicine_en.pdf

[99] Fahim M, Idris M, Ali R, Nugent C, Kang B, Huh EN, Lee S. ATHENA: a personalized platform to promote an active lifestyle and wellbeing based on physical, mental and social health primitives. Sensors (Basel) 2014; 14(5): 9313-9329.

[100] Kim JH. Health avatar: an informatics platform for personal and private big data. Healthc Inform Res 2014; 20 (1): 1-2.

[101] Ashish N, Biswas A, Das S, Nag S, Pratap R. The Abzooba Smart Health Informatics Platform (SHIP): from patient experiences to big data to insights [internet]. arXiv 2012. Available at http://arxiv.org/abs/1203.3764

[102] Lupton D. The commodification of patient opinion: the digital patient experience economy in the age of big data. Sociol Health Illn 2014; 36(6): 856-869.
[103] Cheshire P, Lhotska L, Pharow P. Virtual physiological human and its role for advanced pHealth service provision. In: Blobel B, Pharow P, Parv L, editors. pHealth 2013. IOS Press; 2013.
[104] Ji Z, Ganchev I, O'Droma M, Zhang X, Zhang X. A cloud-based X73 ubiquitous mobile healthcare system: design and implementation. Sci World J 2014; 2014(2014): 1-14.
[105] Jiang P. et al. An intelligent information forwarder for healthcare big data systems with distributed wearable sensors. IEEE Syst J 2014; PP(99): 1-13.
[106] Neill DB. Using artificial intelligence to improve hospital inpatient care. IEEE Intell Syst 2013; 28(2): 92-95.
[107] Ghassemi M, Celi LA, Stone DJ. State of the art review: the data revolution in critical care. Crit Care 2015; 19(118): 801.
[108] Baum S. 4 ways healthcare is putting artificial intelligence, machine learning to use. MedCity News. 2015 [cited 2015 Feb 20]. Available at: http://medcitynews.com/2015/02/4-ways-healthcare-putting-artificial-intelligence-machine-learning-use/
[109] Dixon BE, Rosenman M, Xia Y, Grannis SJ. A vision for the systematic monitoring and improvement of the quality of electronic health data. In: Lehmann CU, Ammenwerth E, Nohr C, editors. MEDINFO 2013. IOS Press; 2013.
[110] Keenan GM. Big data in health care: an urgent mandate to CHANGE nursing EHRs! Online J Nurs Inform 2014; 18(1): 1-3.
[111] Lissovoy G. Big data meets the electronic medical record. Med Care 2013; 51(9): 759-760.
[112] Vishnyakova D, Bottone S, Pasche E, Lovis C. Practical implementation of a bridge between legacy EHR system and a clinical research environment. In: Cross-border challenges in informatics with a focus on disease surveillance and utilising big data, Stoicu-Tivadar L, de Lusignan S, Orel A, Engelbrecht R, Surjan G, eds. Amsterdam, The Netherlands: IOS Press; 2014.
[113] Hoffman S, Podgurski A. The use and misuse of biomedical data: is bigger really better? Am J Law Med 2013; 39(2013): 497-538.
[114] Skaliotis M. Timeliness and accuracy in official statistics 2.0. 2010. Available at https://secure2.gov.mt/NSO/docs/MichailSkaliotis_Eurostat.pdf (accessed 17 August 2014)
[115] European Commission. Communication from the Commission: together for health: a strategic approach for the EU 2008–2013. 2007. COM/2007/630. Available at http://ec.europa.eu/health/ph_overview/Documents/strategy_wp_en.pdf
[116] Phillips KA, Trosman JR, Kelley RK, Pletcher MJ, Douglas MP, Weldon CB. Genomic sequencing: assessing the health care system, policy, and big-data implications. Health Aff 2014; 33(7): 1246-1253.
[117] Kao RR, Haydon DT, Lycett SJ, Murcia PR. Supersize me: how whole-genome sequencing and big data are transforming epidemiology. Trends Microbiol 2014; 22(5): 282-291.
[118] Mavandadi S, Dimitrov S, Feng S, Yu F, Yu R, Sikora U, Ozcan A. Crowd-sourced BioGames: managing the big data problem for next-generation lab-on-a-chip platforms. Lab Chip 2012; 12(20): 4102–4106.
[119] Swan M. Next-generation personal genomic studies: extending social intelligence genomics to cognitive performance genomics in quantified creativity and thinking fast and slow expanded context for personal genomics. In: 2013 AAAI Spring Symposium: 2013 Mar 25-27; Palo Alto, CA, USA. Available at https://www.aaai.org/ocs/index.php/SSS/SSS13/paper/viewFile/5704/5949
[120] 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 2014; 33(7): 1123-1131.
[121] Chute C, Ullman-Cullere M, Wood GM, Lin SM, He M, Pathak J. Some experiences and opportunities for big data in translational research. Genet Med 2013; 15(10): 802-809.
[122] Hofker MH, Fu J, Wijmenga C. The genome revolution and its role in understanding complex diseases. Biochim Biophys Acta 2014; 1842(10): 1889-1895.
[123] Lecroq T, Soualmia LF. From genome sequencing to bedside. Findings from the section on bioinformatics and translational informatics. Yearb Med Inf 2013; 2013(8): 175-177.
[124] Bromley D, Rysavy SJ, Su R, Toofanny RD, Schmidlin T, Daggett V. DIVE: a data intensive visualization engine. Bioinformatics 2014; 30(4): 593-595.
[125] Cole JB, Newman S, Foertter F, Aguilar I, Coffey M. Breeding and genetics symposium: really big data: processing and analysis of very large data sets. J Anim Sci 2012; 90(3): 723-733.
[126] Costa FF. Big data in biomedicine. Drug Discov Today 2014; 19(4): 433-440.
[127] Dinov ID, Petrosyan P, Liu Z, Eggert P, Hobel S, Vespa P, Woo Moon S, Van Horn JD, Franco J, Toga AW. High-throughput neuroimaging-genetics computational infrastructure. Front Neuroinform 2014; 8(41): 1-11.
[128] Hood L. Systems biology and p4 medicine: past, present, and future. Rambam Maimonides Med J 2013; 4(2): 1-15.
[129] Rosenstein BS, West CM, Bentzen SM, Alsner J, Andreassen CN, Azria D, Barnett GC, Baumann M, Burnet N, Chang-Claude J, et al. Radiogenomics: radiobiology enters the era of big data and team science. Int J Radiat Oncol Biol Phys 2014; 89(4): 709-713.
[130] Vivar JC, Pemu P, McPherson R, Ghosh S. Redundancy Control in Pathway Databases (ReCiPa): an application for improving gene-set enrichment analysis in omics studies and "big data" biology. Omics 2013; 17(8): 414-422.
[131] Lueck R. Large-scale genome analysis on helix nebula: the science cloud federated big data processing on-demand. 2014. Available at http://www.helix-nebula.eu/sites/default/files/6. EMBL flagship.pdf (accessed 17 August 2014)
[132] Keen J, Calinescu R, Paige R, Rooksby J. Big data + politics = open data: the case of health care data in England. Policy and Internet 2013; 5(2): 228-243.
[133] Itou T. Big data and official statistics: analysis of recent discussions in statistical communities. 2014. Available at http://www.unescap.org/sites/default/files/2-Big%20Data%20and%20Official%20StatisticsAnalysis%20of%20Recent

%20Discussions%20in%20Statistical%20Communities_-Japan_0.pdf (accessed 17 August 2014)

[134] Baeza-Yates R. Big data or right data? 2013. Available at http://ceur-ws.org/Vol-1087/paper14.pdf (accessed 17 August 2014)

[135] Curtis LH, Brown J, Platt R. Four health data networks illustrate the potential for a shared national multipurpose big data network. Health Aff 2014; 33(7): 1178-1186.

[136] Edwards A, Hollin I, Barry J, Kachnowski S. Barriers to cross-institutional health information exchange: a literature review. J Healthc Inf Manag 2010; 24(3): 22-34.

[137] Elliot T, Holmes JH, Davidson AJ, La Chance PA, Nelson AF, Steiner JF. Data warehouse governance programs in healthcare settings: a literature review and a call to action. eGEMS 2013; 1(1): 1–7.

[138] Karlberg M, Skaliotis M. Big data for official statistics: strategies and some initial European applications. In: Conference of European Statisticians: 2013 Sept 25-27; Geneva, Switzerland. Available at http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2013/mgt1/WP30.pdf

[139] Miller P. Genetic discrimination in the workplace. J Law Med Ethics 1998; 26(3): 189-197.

[140] Riba S. The use of genetic information in health insurance: who will be helped, who will be harmed and possible long-term effects. USC Rev Law Soc Justice 2007; 16(2): 469-489.

[141] Kum HC, Ahalt S. Privacy-by-design: understanding data access models for secondary data. AMIA Jt Summits Transl Sci Proc 2013; 2013(1): 126-130.

[142] European Commission. Communication from the Commission: towards better access to scientific information: boosting the benefits of public investments in research. COM/2012/401. 2012. Available at https://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en.pdf

[143] European Commission. Directive on the application of patients' rights in cross-border healthcare. Directive 2011/24/EU. 2011. Available at http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32011L0024&from=EN

[144] European Commission. Communication from the Commission: on the protection of individuals with regard to the processing of personal data and on the free movement of such data. COM/2012/11. 2012. Available at http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52012PC0011&from=en

[145] World Economic Forum & Boston Consulting Group. Rethinking personal data: strengthening trust. 2012. Available at http://www3.weforum.org/docs/WEF_IT_RethinkingPersonalData_Report_2012.pdf (accessed 17 August 2014)

[146] Megget K. Riding the data stream. PharmaTimes Digital. 2011. Available at http://www.pharmatimes.com/Magazine/Riding_the_data_stream.aspx (accessed 24 January 2014)

[147] Gravelle H, Sutton M, Ma A. Doctor behaviour under a pay for performance contract: treating, cheating and case finding? Econ J 2010; 120(542): F129-F156.

[148] Sigfrid LA, Turner C, Crook D, Ray S. Using the UK primary care Quality and Outcomes Framework to audit health care equity: preliminary data on diabetes management. J Public Health (Oxf) 2006; 28(3): 221-225.

[149] Millett C, Netuveli G, Saxena S, Majeed A. Impact of pay for performance on ethnic disparities in intermediate outcomes for diabetes: a longitudinal study. Diabetes Care 2009; 32(3): 404-409.

[150] European Collaboration for Healthcare Optimisation (ECHO). What is the ECHO project? 2014. Available at http://echo-health.eu/ (accessed 24 January 2015)

[151] NHS England. The care.data programme: better information means better care. Technology, systems and data. 2014. Available at http://www.england.nhs.uk/ourwork/tsd/caredata/ (accessed 17 August 2014)

[152] Parliamentary Office of Science and Technology. Transparency of clinical trial data. 2014. Available at http://www.parliament.uk/briefing-papers/post-pn-461.pdf (accessed 24 January 2015)

[153] Krumholz HM, Ross JS, Gross CP, Emanuel EJ, Hodshon B, Ritchie JD, Low JB, Lehman R. A historic moment for open science: the Yale University Open Data Access project and medtronic. Ann Intern Med 2013; 158(12): 910-911.

[154] PMLiVE. Pharma industry announces data sharing commitment. 2013. Available at http://www.pmlive.com/pharma_news/pharma_industry_announces_data_sharing_commitment_491944 (accessed 24 January 2015)

[155] Institute of Medicine. Sharing clinical trial data: maximizing benefits, minimizing risk. 2015. Available at http://www.iom.edu/Reports/2015/Sharing-Clinical-Trial-Data.aspx (accessed 24 January 2015)

[156] World Health Organization. WHO statement on public disclosure of clinical trial results. 2015. Available at http://www.who.int/ictrp/results/Draft_WHO_Statement_results_reporting_clinical_trials.pdf (accessed 24 January 2015)

[157] Audit Commission. Data remember: improving the quality of patient based information in the NHS. 2002. Available at http://archive.audit-commission.gov.uk/auditcommission/sitecollectiondocuments/AuditCommissionReports/NationalStudies/dataremember.pdf (accessed 24 January 2015)

[158] Audit Commission. Data assurance framework. 2012. Available at http://www.audit-commission.gov.uk/information-and-analysis/data-assurance-framework/ (accessed 24 January 2015)

[159] Spencer A. Hospital episode statistics (HES): improving the quality and value of hospital data. 2011. Available at http://www.hscic.gov.uk/media/1593/Hospital-Episode-Statistics-Improving-the-quality-and-value-of-hospital-data/pdf/HES_-_Improving_the_quality_and_value_of_hospital_data.pdf (accessed 24 January 2015)

[160] Frisse M, Wilcox A, Sittig D, Kahn M, Lopez MH. Clinical informatics, CER, and PCOR: building blocks for meaningful use of big data in health care. In: Medical Care Special Supplement Webinar Series (Electronic Data Methods) Forum: 2012 Oct 31; Webinar. Available at http://repository.edm-forum.org/webinars/6/

[161] Rial N. The power of big data in Europe. New Europe. 2013. Available at http://www.neurope.eu/article/power-big-data-europe (accessed 17 August 2014)

[162] European Commission. Improving health reporting mechanisms. Public health. 2015. Available at http://ec.europa.eu/health/data_collection/tools/mechanisms/index_en.htm (accessed 24 January 2015)

[163] European Health Forum Gastein. Conference report: resilient and innovative health systems for Europe. 2013. Available at http://www.ehfg.org/fileadmin/ehfg/Programm/2013/EHFG_2013_Conference_Report.pdf (accessed 24 January 2015)

[164] European Commission. Press release: better use of health data will transform the healthcare landscape, says expert report. 2012. IP/12/453. Available at http://europa.eu/rapid/press-release_IP-12-453_en.pdf

[165] European Commission. Communication from the Commission: a comprehensive approach on personal data protection in the European Union. 2010. COM/2010/609. Available at http://ec.europa.eu/justice/news/consulting_public/0006/com_2010_609_en.pdf

[166] Diverty B. Drowning in data? Health analytics, a lifejacket for performance measurement. 2014. Available at http://www.longwoods.com/articles/images/Brent Diverty Deck.pdf (accessed 17 August 2014)

[167] Kaplan RM, Chambers DA, Glasgow RE. Big data and large sample size: a cautionary note on the potential for bias. Clin Transl Sci 2014; 7(4): 342-346.

[168] Lamarche-Vadel A, Pavillon G, Aouba A, Johansson LA, Meyer L, Jougla E, Rey G. Automated comparison of last hospital main diagnosis and underlying cause of death ICD10 codes, France, 2008–2009. BMC Med Inform Decis Mak 2014; 14(44): 1-9.

[169] Harron K, Wade A, Gilbert R, Muller-Pebody B, Goldstein H. Evaluating bias due to data linkage error in electronic healthcare records. BMC Med Res Methodol 2014; 14(36): 1-10.

[170] European Parliament & The Council. Decision: establishing a second programme of community action in the field of health. 2007. Decision No 1350/2007/EC. Available at http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32007D1350&from=EN

[171] European Commission. European core health indicators (ECHI). Public health. 2015. Available at http://ec.europa.eu/health/indicators/echi/index_en.htm (accessed 24 January 2015)

[172] Accenture. Global market for electronic health records (EHR) expected to reach $22.3 billion by the end of 2015. 2014. Available at http://newsroom.accenture.com/news/global-market-for-electronic-health-records-expected-to-reach-22-3-billion-by-the-end-of-2015-according-to-accenture.htm (accessed 24 January 2015)

[173] Todd R. E-Health Insider: EMIS and TPP share data. ehi News. 2013. Available at http://www.ehi.co.uk/news/EHI/8458/emis-and-tpp-share-data (accessed 24 January 2015)

[174] Berndt DJ, Fisher JW, Hevner AR, Studnicki J. Healthcare data warehousing and quality assurance. Computer (Long Beach Calif) 2001; 34(12): 56-65.

[175] Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, Riley WT, Shar A, Spring B, Spruijt-Metz D, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013; 45(2): 228-236.

[176] Kayyali B, Knott D, Kuiken S Van. The big-data revolution in US health care: accelerating value and innovation. 2013. Available at http://www.mckinsey.com/insights/health_systems_and_services/the_big-data_revolution_in_us_health_care (accessed 17 August 2014)

[177] Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH. Big data: the next frontier for innovation, competition, and productivity. 2011. Available at http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation (accessed 24 January 2015)

[178] Hervey T, Vanhercke B. Health care and the EU: the law and policy patchwork. In: Mossialos E, Permanand G, Baeten R, Herv TK, editors. Health systems governance in Europe: the role of EU law and policy. New York: Cambridge University Press; 2010.