DOSSIER “PRIVACY: DEFINITION, PROTECTION AND PROJECTION”

Thoughts

Legal protections for personal health information in the age of Big Data — a proposal for regulatory framework

Protection légale des données de santé personnelles à l’heure du Big Data – une proposition pour un cadre réglementaire

D. Mendelson (Research Chair in Law School, Key Researcher)\textsuperscript{a,b,*}, D. Mendelson (Consultant Engineer)\textsuperscript{c}

\textsuperscript{a} Deakin University, Melbourne, Australia
\textsuperscript{b} Law and Policy Program, Data to Decisions Cooperative Research Centre, Melbourne, Australia
\textsuperscript{c} SYPAQ Systems Pty Ltd, Melbourne, Australia

Received 25 November 2016; accepted 24 January 2017
Available online 18 March 2017

KEYWORDS
Big Data; Health data; Privacy; Protection; Telemedicine

Summary This article outlines the need for, and possible solutions to, the problem of third parties’ legally untrammeled ability to collect and use identified or identifiable personal, sometimes very sensitive, health data. A solution would be implementation of a comprehensive legal framework — from international treaties through national legislation to operationalising data privacy and ethics by design at the level computer software (algorithmic) instructions. At least since the last decade of the twentieth century, digitisation of health data and creation of national electronic health records infrastructure has held the promise of enabling the attainment of such public health goals as personal health management, health care delivery, health-related research, and population health surveillance. Great advances in Big Data technology, and even more so, the algorithmic revolution, has facilitated these four goals, though not necessarily in the ways envisaged by scholars and policy-makers of the time. Thus, personal health management is supported by apps such as Apple Health; telemedicine and teleradiology systems enable health care delivery to patients wherever they are located. Health-related

\textsuperscript{*} We are grateful to Dr Gabrielle Wolf, Dr Elizabeth Haywood and Prof David Weisstub for their very helpful comments.
\textsuperscript{*} Corresponding author. Deakin University, Burwood Campus, Building EA, Elgar Rd Burwood, VIC 3125, Melbourne, Australia.
E-mail address: danuta.mendelson@deakin.edu.au (D. Mendelson).

http://dx.doi.org/10.1016/j.jemep.2017.02.005
2352-9525/© 2017 Elsevier Masson SAS. All rights reserved.
medical, commercial, socio-economic and socio-political research based on Big Data is booming, while national electronic health record systems allow national and international agencies to track and scrutinize health of individuals and populations. However, unregulated and rampant “datafication” of identified or identifiable personal health information about individuals collected, managed, and disseminated without their knowledge and informed consent effectively treats data subjects — us — as mere means to an end. The law has been lagging a long way behind technical and commercial development, yet it is possible to safeguard privacy and other fundamental rights of data subjects. For example, as part of the European Union’s Digital Single Market Strategy, the European Parliament adopted Regulation (EU) 2016/679 on the “protection of natural persons with regard to the processing of personal data and on the free movement of such data”, and the Directive (EU) 2016/680 on the “protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data”. The Regulation shall apply, and the Directive will enter into force (requiring EU Member States to transpose it into their national law) in May 2018. In the United Kingdom, the Investigatory Powers Act (UK) 2016 has systemically incorporated approach to statutory controls in form of proportionality and necessity tests on powers that national security and law enforcement agencies must intercept, communications, access, collect and manage massive volumes of data (known as “bulk powers”). These controls aim to create privacy safeguards for intentionally or inadvertently targeted individuals. Thus, the law seems to be “awakening”. However, a comprehensive and systematic regulatory framework of controls and protections is yet to be postulated. This article outlines an approach consisting of vertical tiers that can be implemented separately or in total. The article has two parts. The first part provides background to the interface between developments in technology and unconsented to “datafication” of our personal health from which, unknown to us, third party algorithms create our digital identity. The second part outlines proposal that envisages a five-tier legal framework for protection of identifiable personal health data. Protections at each tier would be discrete yet capable of integration. At the base — the design level — it possible for software engineers and computer programmers to specify precisely defined algorithmic instructions for processing personal data in accordance with privacy laws and ethical standards. At the next level — data operators and/or analysts (whether human or an automated algorithmic program) — legal tests of proportionality and necessity and ethical conduct could be implemented by legislation and embedded in the algorithm design. At the third level — individuals and organizations storing and using the data — responsibility for its integrity and security as well as privacy and ethics could be governed by legislation. At the top level, international treaties could provide for uniform standards and approach.

© 2017 Elsevier Masson SAS. All rights reserved.

MOTS CLÉS
Big Data ;
Données de santé ;
Intimité ;
Protection ;
Télémédecine

Résumé Cet article étudie le problème et propose des pistes de solutions au problème de tierces personnes ayant la capacité, légalement incontrôlable, d’accumuler et d’utiliser des données identifiées ou identifiables, parfois très confidentielles, sur la santé. Une des solutions qui s’offre serait la mise en place d’un cadre réglementaire exhaustif – des traités internationaux, en passant par la législation nationale, puis à la mise en place d’un système de données confidentielles et d’éthique à partir d’un logiciel d’instruction (algorithme). Depuis au moins la dernière décennie du vingtième siècle, la digitalisation des données sur la santé et la création de dossiers de santé en ligne à l’échelle nationale permettent d’atteindre les objectifs de santé publique tels que la gestion de la santé personnelle, la prestation de soins, la recherche en santé et la surveillance de la santé de la population. Les avancées dans la technologie des mégadonnées, et qui plus est, la révolution algorithmique, a facilité l’accomplissement de ces quatre objectifs, qu’elle d’une façon différente à celle envisagée par les chercheurs et les responsables politiques de l’époque. En effet, la gestion de la santé personnelle est encouragée par des applications comme Apple Health ; des systèmes de télémédecine et de téléradiologie permettent la livraison de soins aux patients n’importe où. Les recherches médicale, commerciale, socioculturelle et sociopolitique liées à la santé et basées sur les mégadonnées sont en pleine essor, alors que des systèmes nationaux de dossiers de la santé électroniques permettent aux agences nationales et internationales de tracer et d’examiner la santé des individus et des populations. Cependant, la « datafication » non réglementée et rampante d’information
personnelle sur la santé des individus qui est accumulée, gérée et diffusée sans qu'ils ne le sachent et sans leur consentement traiter les sujets de ces données — nous — comme le simple moyen une finalité autre. Le droit a pris du retard face aux développements techniques et commerciaux, mais il demeure possible de protéger la confidentialité et autres droits fondamentaux des sujets de ces données. À titre d'exemple, dans le cadre de la stratégie pour un marché unique numérique de l'Union européenne, le Parlement européen a adopté le règlement (UE) 2016/679 relatif à la « protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données » et la directive (UE) 2016/680 relative à la « protection des personnes physiques à l'égard du traitement des données à caractère personnel par les autorités compétentes à des fins de prévention et de détection des infractions pénales, d'enquêtes et de poursuites en matière où d'exécution de sanctions pénales, et à la libre circulation de ces données ». Le règlement s'applique et la directive entrera en vigueur (lorsque les membres de l'UE auront transposé cette loi dans leur loi nationale) en mai 2018. Au Royaume-Uni, l'Investigatory Powers Act (UK) 2016 a systématiquement incorporé une approche à ce contrôle statutaire par l'entremise de tests de proportionnalité et de nécessité sur les pouvoirs conférés à la sécurité nationale en matière d'interception, de communication, d'accès, de collection et de gestion de quantité massive de données (connus sous le nom de « Bulk powers »). Ce contrôle a pour but de créer des mesures appropriées pour la protection de la vie privée pour les individus qui sont intentionnellement ou involontairement ciblés. Ainsi, le droit semble s'éveiller. Toutefois, un cadre réglementaire exhaustif et systématique de contrôle et de protection n'a pas encore été instauré. Cet article propose une approche en étagées verticales qui pourraient être mises en œuvre séparément ou ensemble. Cet article est divisé en deux parties. La première partie donne de l'information sur l'interface entre les développements en technologie et notre non-consentement à la « datafication » de notre santé personnelle, à partir de laquelle, à notre insu, les algorithmes de tiers parties créent notre identité digitale. La deuxième partie émet une proposition qui envisage un cadre légal en quatre temps pour la protection des données identifiables sur la santé. Les protections à chaque étage seraient discrètes mais capables d'intégration. À la base — au niveau de la conception — il est possible pour les ingénieurs informatiques et les programmeurs informatiques de spécifier des algorithmes d'instructions définis de façon précise pour traiter les données personnelles conformément aux lois sur la confidentialité et les standards éthiques. Au niveau suivant — les opérateurs de données et les analystes (humains ou un programme algorithmique automatisé) — des tests légaux de proportionnalité et de nécessité et de conduite éthique pourraient être mis en place par la législation et intégrés dans la conception algorithmique. Au troisième niveau — les individus et organisations qui stockent et utilisent — la responsabilité de l'intégrité, la sécurité ainsi que la confidentialité et l'éthique pourraient être gouvernées par la législation. Et au plus haut niveau, les traités internationaux pourraient établir des standards et une approche uniforme.

© 2017 Elsevier Masson SAS. Tous droits réservés.

This article outlines the problem of third parties’ legally untrammelled ability to collect and use identified or identifiable personal, sometimes very sensitive, health data — and possible solutions to this problem. The proposed response is to implement a comprehensive legal framework: involving international treaties, national legislation, and, in particular, the mandatory incorporation of data privacy and ethics principles within the design of computer software.

Since the last decade of the twentieth century, the digitisation of health data and creation of national electronic health records infrastructure has held the promise of transforming fundamental pillars of public health such as personal health management, health care delivery, health-related research, and population health surveillance. Great advances in Big Data technology (e.g., predictive analytics, automated reporting of patient data) have facilitated the achievement of these transformations, though not necessarily in ways envisaged by scholars and policy-makers of the time.

Personal health management is supported by apps such as Apple Health, and telemedicine and teleradiology systems enable health care delivery to patients wherever they are located. Health-related medical, commercial, socioeconomic and socio-political research based on Big Data is booming, while national electronic health record systems allow national and international agencies to track and scrutinise the health of individuals and populations. However, as the authors of the 2014 Report on “Big Data: Seizing Opportunities, Preserving Values” [2] noted “these capabilities, most of which are not visible or available to the average consumer, also create an asymmetry of power between those who hold the data and those who intentionally or inadvertently supply it”.

A consequence of this asymmetry is that rampant “datafication” of identified or identifiable personal health information about individuals is being conducted with virtually no regulation and oversight. The term “datafication” does not have a consistent definition [2] but for the purposes
of this article will be defined as: the collation and storage of data inherent in, or related to, objects, processes or people; and the transformation of this data into new forms, which possess value related — or unrelated — to the data's original context.

Datafication takes processes and data sources previously "invisible" and analyses them to create valuable information. Every phone call, e-mail, financial transaction, transportation ticket or stroll past a security camera creates data, which adds to various pools available to governments and private companies. For example, it is claimed that datafication enables accurate prediction of "personality types using mobile phone data" (i.e. analysing "mobile phone logs to determine the way in which a person uses the apps they have downloaded can provide a wealth of different information, such as credit rating, risk profile, trustworthiness indicators and more") [3].

Profound legal and ethical questions arise when health insurance companies exploit Big Data to identify risky insureds. For example, Blue Cross Blue Shield of Tennessee (BCBS) in cooperation with a Big Data in-database analytics firm Fuzzy Logix [4], is using "years" worth of pharmacy data, along with claims data, both from BCBS's customers and others, through a third-party sharing agreement [5] to highlight "with something like 85% accuracy" [6], "risk factors that could indicate whether a person could be in danger of developing an opioid abuse problem." [6]. According to the BCBS, the major reasons for using data mining and predictive analytics on healthcare databases are "to allow doctors and other specialists to step in and offer preventative care before the [predicted] problem gets out of hand", and to reduce insurance expenses. Apparently in the US, opioid drugs "abusers are 59% more likely to be high-cost claimants, with healthcare costs of greater than $50,000 per year. By rolling out predictive, preventative treatment nationwide, in theory insurance costs should come down for everyone" [6]. As part of this process, individuals identified as having the potential, or a "chance" [6], of developing opioid drug addiction are being tagged. These tags then flow to other integrated or aggregated data sets. In-database techniques mean that the analytics footprint is very light; it can be performed on large data sets anywhere in the world, and may not be reportable.

This information is being processed (collected, managed, analysed and disseminated) without the knowledge and informed consent of the people to whom it relates. Consequently, we — the data subjects — are being treated as a fertile source of data to harvest, towards commercial or government ends determined by the data processors. Although the law has lagged a long way behind technical and commercial developments, it may be possible to go some way towards safeguarding privacy and other rights of data subjects. It is not feasible to impose a world-wide, all-embracing legal harness of privacy regulation over the ever-expanding universe of data. Nonetheless, it is vital to discuss legal measures that may enable the implementation of systemic privacy safeguards for healthcare data collection and use.

Thus, as part of the European Union's Digital Single Market Strategy, the European Parliament adopted Regulation (EU) 2016/679 on the "protection of natural persons with regard to the processing of personal data and on the free movement of such data", and Directive (EU) 2016/680 on the "protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data". The Regulation shall apply, and the Directive will enter into force (requiring EU Member States to transpose it into their national law) in May 2018. The Regulation and the Directive set up a very comprehensive, meticulous, and strict legal framework that may not be appropriate for the world of fast changing technology [7].

In the US, the government has opted for the approach that targets specific sectors involved with Big Data and the Internet of things, referred to as IoT [8], focusing on particular problems arising from their development and usage [9]. Regulation [10] accompanied by voluntary codes are the preferred tools of enforcement; the rules and standards are product of consultations with the regulators (the Congress and the Senate), technology business and industry representatives, National Academies of Science, Engineering, and Medicine, as well as academics. For instance, the Federal Trade Commission not only exercises its powers to prosecute law violations of consumers' privacy and security, but also conducts studies and issues reports regarding legislative and regulatory proposals that affect consumer privacy [11]. Another example is the "Precision Medicine Initiative: Privacy and Trust Principles" [12] developed by the White House Office of Science and Technology Policy, the Department of Health and Human Services Office for Civil Rights, and the National Institutes of Health to govern data generated in the context of "precision Medicine" practice (tailoring medical treatment options to the individual patient's "genome sequence, microbiome composition, health history, lifestyle, and diet") [13]. Though less systemic, this approach provides for a more dynamic relationship with the relevant technological innovations, though at the cost of delay in implementation.

In the United Kingdom, the Investigative Powers Act 2016 (UK) has systematically incorporated statutory controls that require the application of proportionality and necessity tests before national security and law enforcement agencies' use their powers to intercept communications, and to access, collect and manage massive volumes of data (known as "bulk powers"). These controls aim to create privacy safeguards for intentionally or inadvertently targeted individuals. Thus, the law seems to be "awakening" to the challenges to privacy posed by potential government misuse of bulk data. However, a comprehensive and systematic regulatory framework of controls and protections is yet to be postulated. This article outlines an approach consisting of vertical tiers of regulation that can be implemented separately or in toto.

This article has three parts. The first, introductory part 1 outlines our concept of a five-tiered inverted triangle (Fig. 1) that illustrates what happens to the digitised personal health-related information we generate. The first part provides background to the interface between developments in technology and "datafication" of our personal health information from which, without our knowledge and consent, third party processors create our digital identities. The second part discusses the European, UK and US models
for the protection of identifiable personal health data. The third part proposes a framework based on tests of proportionality, necessity and balance for the incorporation of legal privacy protections at each of the layers of the inverted triangle. Tests proposed for each tier are discrete, yet capable of integration.

**Part 1**

It is a truism to write that advances in technology are instrumental in shaping and in re-shaping societies. How can a society in the process of re-shaping itself safeguard the rights of individuals, the social values, and the ethical principles that have historically underpinned that society? Can these rights, values and principles be adapted to the technologies and structures of the new era?

Before responding to these questions, it is apposite to provide a brief background to transformative technologies. Historically, physical artefacts created by new technologies are visible, be it the wheel, plough, steam engine, telephone, aeroplane or antibiotics. Personal computers were symbols of the World Wide Web (invented in 1989 by Tim Berners-Lee)\(^1\), which in 1991 ushered in the Internet-based "networked digital age". Since then, the convergence of many new technologies, in particular, "Big

\(^1\) Developments leading to personal computers and the WWW trace back to the 1940s.
Data\textsuperscript{2} \cite{14} and the IoT, has led to what is sometimes called the digital transformation era.

However, the essential artefacts of the new era — digitised data and algorithms to process that data — are ubiquitous yet virtually invisible. Hence we may not be fully aware that, among others, "datafication" (which is rendering "into [electronic] data many aspects of the world that have never been quantified before") \cite{15}, and automated algorithms are transforming the way we organise our lives; conduct research; provide healthcare and commercial and government services.

Nonetheless, awareness of technological drivers is essential to the critical assessment of both the great benefits and significant drawbacks that this revolution is creating. In respect of governance, one of the most important of these drawbacks is either over-regulation or insufficient regulation.

We focus on the health domain, and argue that legal principles protecting rights of individuals, in particular, tests of proportionality and necessity, should underpin instruments of legal regulation requiring the use of engineering standards to embed privacy controls into the design of systems and software, and that laws are needed to impose a measure of legal responsibility for the security and integrity of personal data on all organisations that collect and manage it, including third parties.

We conceptualise the "datafication" of personal health-related information as an inverted triangle with five layers. Below the sharp end of the triangle, we — each of us — generate health-related information in identified form and input it, either intentionally or inadvertently, into the triangle.

The first layer comprises our ready-to-be-datafied personal information as collected and processed by smartphones, internet searches, purchases, medical and fitness devices, our healthcare providers (such as pharmacies, health care clinics and hospitals), and through electronic health records.

The second layer consists of clinical and other health-related data in identified form that is collected, stored and distributed to third parties (medical facilities, healthcare insurance organisations, business analytics and marketing analysis firms, as well as public agencies that may not be health-related, for example, law enforcement and national security bodies).

The third layer is closely interwoven with the second layer, and involves data broker mechanisms that are part of high performance distributed computing environments' architecture \cite{16}. Mainly private companies, called data brokers or data-aggregators, unbeknown to data subjects, collect raw data for analysis from the two lower layers of the inverted triangle, as well as from other public and private sources that contain personal, often health-related information about us generated by others. As the example of Blue Cross Blue Shield of Tennessee health insurance fund and Fuzzy Logic illustrates, the integration and data mining of previously independent datasets "can lead to surprising or unintended inferences across these newly integrated datasets, resulting in previously unknown [accurate or otherwise] facts about subjects" \cite{17}. The datafied/processed results are then sold or distributed either in an identifiable or de-identified form \cite{18}.

The fourth layer includes national governments, and international public and private entities, that use, re-sell, re-distribute and re-process our personal health-related data for various purposes (marketing products, detecting fraud \cite{19}, research\textsuperscript{3} \cite{20}, government policy planning and decisions \cite{21}, and surveillance \cite{22}).

The fifth layer — still theoretical — would comprise of international treaties and agreements governing privacy protections for personal health-related data. The Global Privacy Enforcement Network (GPEN) was established in 2010, "to promote and support cooperation in cross-border enforcement of laws protecting privacy" among its 27 participating privacy enforcement "authorities" \cite{23}. However, the GPEN, or any other international organization, is yet to produce a blueprint for cross-border privacy laws.

This article is premised on the assumption that current efforts by the international software engineering community to develop standards and professional guidance for "Privacy Engineering" are formalised, accepted and widely adopted internationally. At the present time, the National Institute of Standards and Technology (NIST, part of the U.S. Department of Commerce) has published an internal report "containing a general roadmap for evolving [Privacy Engineering] preliminary concepts into actionable guidance" \cite{24}. The NIST "Privacy Control Catalog" \cite{25} provides high-level guidance at the system and design level for the implementation of privacy controls to improve privacy. These only apply to the US Federal Government\textsuperscript{4}, not to private industry or to US State Governments.

Such controls include organisational processes (for example, continual review of the need to store personally identifiable information), software features (for example, accurate account of each time a personal record is disclosed and making that accounting of disclosures available to the person named in the record upon request), and guidance for the use of personal data (for example, minimising the use of "real"

\textsuperscript{3} This study "integrated multiple nationwide administrative databases [in New Zealand] and electronic medical records" with the four-decade-long "Dunedin Longitudinal Study, a population-representative 1972–1973 birth cohort of 1037 New Zealanders assessed at ages 3, 5, 7, 9, 11, 13, 15, 18, 21, 26, 32 and 38 years and followed from birth to midlife with 95% retention". The analysis tested how strongly childhood risks predicted "social welfare dependency, fatherless child-rearing, tobacco smoking, excess body weight, admissions to taxpayer-funded national-health service hospitals, taxpayer-funded prescription drug fills, taxpayer-funded insurance claims paid out for accidents and injuries, and convictions for crime".

\textsuperscript{4} For example, insureds may be asked by a health insurance organisation to "voluntarily" undergo low cost DNA tests that predict their future health conditions.

\textsuperscript{5} For example, the Asia Pacific Economic Cooperation, the International Conference of Data Protection and Privacy Commissioners, the EU Article 29 Working Party, and the Organization for Economic Cooperation and Development.

\textsuperscript{6} The US Office of Management and Budget, 2016 update to Circular No. A-130 requires US Federal agencies to apply the NIST Risk Management Framework (RMF) in their privacy programs.

\textsuperscript{2} The term "data revolution" was coined in 2013.
data in software testing and in organisational training). Privacy Engineering is not yet formalised in internationally-recognised standards published by the Institute of Electrical and Electronics Engineers (IEEE) or the International Organization for Standardization (ISO); however, some large companies are already beginning to consider how to incorporate privacy protections into their engineering processes [26].

Additionally, this article is premised on the expectation that data operators and/or analysts (whether human or automated algorithmic programs) are or will be able to apply legal tests of privacy-oriented proportionality and necessity, as well as codes of ethical conduct that derive from voluntary codes and legislation. At the third layer, data brokers could be governed by legislation imposing requirements of transparency, integrity, as well as compliance with privacy-oriented proportionality tests. At the fourth layer, tests modelled on the Investigative Powers Act 2016 (UK) could be adapted to apply to all governmental agencies and instrumentalities, while at the top-fifth layer, international treaties may provide for uniform standards and approaches [27].

Background

A major aspect of data transformation has been "Big Data", a catch-all term that encompasses technological changes enabling data (including identified personal health data) to be accessed, collected, and aggregated from diverse sources in varying formats that are stored, linked, networked, examined and manipulated across different contexts in enormous volume and speed [28]. For example, whereas in 2000, three quarters of all "the world's stored information" was in non-digital form (paper, film, and other analog media), by 2013, "less than two percent of all stored information...was non-digital" [29].

The volume, speed, diversity of sources (e.g. public government records, including digitized healthcare records, internet searches and transactions, social media, mobile phones and the IoT devices), as well as storage of vast amounts of data have been made possible by developments in hardware systems. Almost instantaneous automated aggregation, linkage, examination, and processing (filtering, classifying, drawing inferences) and distribution of data has been enabled by the "algorithmic revolution" [30] -- the science of data analytics based on novel and even more powerful complex mathematical, computational, and statistical algorithms [31]; digital signals processing algorithms; and autonomous algorithms [32]. Machine learning [33], including "deep learning" [34] algorithms, identify and reveal knowledge about individuals through "aggregation of incongruent data" [28], enabling discovery of "correlations that would otherwise go unnoticed" [28], patterns and predictions about behaviour [35,36] and health conditions of individuals [37,38], groups and communities [28]. These outputs and interpretations are relied on by governments and agencies, banks, businesses, insurance companies, and healthcare professionals for medical diagnostics, marketing, designing policies and making operational decisions at all levels of government and business [39].

The partnership between Blue Cross Blue Shield of Tennessee and Fuzzy Logix is not unique. A high proportion of the data collected and analysed by algorithms includes personal, frequently sensitive, medical or health-related information contained in electronic health records [40]. Thus, Crossix Solutions, a United States healthcare analytics firm has a "proprietary network of health and non-health data covering over 250 million U.S. consumers (76% of the U.S. population)" [41], that include in addition to prescription purchase records (Rx), "hospital records, electronic health records (EHR) and electronic medical records (EMR), doctors’ notes, lab results, and other clinical data" [42]. According to Jeremy Mittler, VP, Industry Solutions at Crossix Solutions, the company can:

"link, for example, the information gleaned from doctor notes to bloodwork results to Rx usage data to individuals exposed to display or mobile ads [which] offers a veritable wealth of insight into what factors trigger certain [marketing] actions for distinct patient segments at different phases of their disease progression" [42].

Health and medical information is also posted on and collected from social media (Facebook, Twitter); Internet searches (Google, Bing, etc.), online purchases, as well as fitness and medical devices. The volume and variety of medical and health information has escalated exponentially since the advent of wireless mobile sensor networks [43] in the IoT, whereby sensors, actuators, computational intelligence and telecommunications systems embedded in everything from cars through to smart phones [44], wearables [45] and medical implants [46] track personal activities, monitor "your biosignals and distill them into meaningful insights."

---

7 An algorithm involves "a precisely defined sequence of instructions for carrying out a particular task. When that task is performed by a computer, the algorithm must be expressed according to the strict rules of a programming language that is ‘understood’ by the computer."

8 "‘Modern machine learning is a statistical process that starts with a body of data and tries to derive a rule or procedure that explains the data or can predict future data. . . . machine learning relies on statistical methods to find a decision procedure that works well in practice.”

9 "Deep learning algorithms use ‘structures loosely inspired by the human brain, consisting of a set of units (or ‘neurons’). Each unit combines a set of input values to produce an output value, which in turn is passed on to other neurons downstream. For example, in an image recognition application, a first layer of units might combine the raw data of the image to recognize simple patterns in the image; a second layer of units might combine the results of the first layer to recognize patterns of patterns; a third layer might combine the results of the second layer; and so on... sometimes more than 100 [layers]... use a large number of units at each layer, to enable the recognition of extremely complex, precise patterns in data.”

10 "Wireless mobile sensor networks ‘are composed of sensor nodes having capability of sensing environmental conditions, sink node [base station] and connected via internet to remote controller’. They are used for environmental monitoring, surveillance, national security and health care."
obtain our health data directly from us. Unaware of which organisation, who, when and how uses our health information, we are unable to verify that this data is genuine and accurate [20]. Yet any erroneous information may affect the verification of our identity by health insurance companies, banks, law enforcement and security agencies, and adversely determine our credit rating, employment, insurance, housing, and the like.

Data broker entities are a relatively new phenomenon — a corollary of the algorithmic revolution; hence the public is still unfamiliar with what they are and the information sources they use. These data miners and advanced analytics entities together with second level data providers could be regulated through imposition of statutory duties to incorporate privacy protection mechanisms into the collection and analytical software for the benefit of patients as a discrete class of the public.\(^{12}\)

Part 2 — privacy protection principles

In the unregulated world of third parties being able to collect, manipulate, and disseminate identified or identifiable personal health information about us without our knowledge and informed consent, we — the data subjects — are treated as a mere means to an end. Ethical questions regarding the commodification of personal identity [55] are outside the scope of this article; however, a comprehensive system of legal frameworks at national and international levels can provide a degree of protection for our privacy, and at least some of our civil rights.

Most legal privacy protections for health-related data were drafted in the twentieth century for technology available at that time [56], and are outdated in the era of Big Data. However, by the mid-1990s Ann Cavoukian, the Information and Privacy Commissioner of Ontario, Canada and John Borking of the Dutch Data Protection Authority and the Netherlands Organisation for Applied Scientific Research co-authored a report on "Privacy-enhancing technologies: the path to anonymity" (1995). The report introduced the concept of "privacy by design", which Ann Cavoukian, writing in 2011, defined as extending to three "encompassing applications: (1) IT systems; (2) accountable business practices; and (3) physical design and networked infrastructure" [57]. She proposed the following seven foundational "privacy by design" principles:

1. Proactive not Reactive; Preventative not Remedial (aiming to anticipate and prevent privacy risks and privacy infractions from materialising);
2. Privacy as the Default Setting (building into IT systems and business practices privacy protections by default);
3. Privacy Embedded into Design (embedding privacy safeguards into the design and architecture of IT systems and business practices thus making it "an essential component of the core functionality being delivered");
4. Full Functionality — Positive-Sum, not Zero-Sum (accommodating all legitimate interests and objectives, for example security, in a positive-sum "win-win" manner);

\(^{11}\) The WoT [Web of Trust] app, available on Chrome and Firefox, was supposed to remove any identifying details from the data it collects (account names, mailing addresses and browsing history) about more than 140 millions of its users. However, German television channel NDR found that the anonymised data could be matched to individual users, and "with access to browser history, NDR could determine users' travel plans, shopping habits, general medical histories and even sexual preferences". Douglas Bonderud, "WoT Privacy Breach: Trust Tanks as Browser Add-On Caught Selling User Data" SecurityIntelligence (10 November 2016) https://securityintelligence.com/news/wot-privacy-breach-trust-tanks-as-browser-add-on-caught-selling-user-data/

\(^{12}\) X (Minors) v Bedfordshire County Council [1995] 2 AC 633.
5. End-to-End Security – Full Lifecycle Protection (having embedded privacy protections into the system prior to the first element of information being collected, extending these protections throughout the entire lifecycle of the data involved);

6. Visibility and Transparency – Keep it Open (assuring all stakeholders through independent verification that the relevant business practices or technologies are "operating according to the stated promises and objectives");

7. Respect for User Privacy – Keep it User-Centric (ensuring that "architects and operators … keep the interests of the individual uppermost by offering such measures as strong privacy defaults, appropriate notice, and user-friendly options") [57].

Principles of "privacy by design and by default" have influenced two different models and approaches to controlling third parties' ability to collect, process and use identified or identifiable personal data. The first model is the General Data Protection Regulation (EU) 2016/679 (GDPR), which adopts a detailed (it contains 173 Recitals and 99 Articles with copious sub-clauses), centralised and highly prescriptive methodology. In contrast, the US approach provides an example of a de-centralised regulation model. US regulatory agencies draft privacy regulations in consultation with business and industry. This model is dynamic but relatively slow - it relies on identifying principles and standards for protection of security, privacy, and trust "when new cyber-related technologies and policies are conceived" [58] on a sector by sector basis, and then testing them in cooperation with the relevant industry segment.

Although the aim of both the EU and US models is to safeguard data subjects' rights to data privacy, they tend to emphasise somewhat different rights and obligations, and therefore focus on different layers of the inverted triangle (refer back to Fig. 1). The two models, as they pertain to health-related data, will be discussed in turn.

**European Union’s general data protection regulation (EU) 2016/679 (GDPR)**

The GDPR [59] was enacted on 27 April 2016 as a result of tri-partite negotiations among the European Parliament, the European Council and the European Commission. It provides a top-down regulatory framework that "shall be binding in its entirety and directly applicable in all Member States" of the EU from 25 May 2018. The strict and detailed code aims to harmonise legal rights of natural persons (below the sharp end of the inverted triangle) for the protection of personal data, including health data [60] "in respect of processing activities" [61]. These activities extend to "the processing of personal data wholly or partly by automated means" [62], thus encompassing all digital data. The GDPR targets the first, the third, and partially the fourth, layers of the triangle. However, Article 23 of the GDPR allows the EU and the Member States to restrict the scope of the codified obligations and rights in relation to data protection "when such a restriction respects the essence of the fundamental rights and freedoms and is a necessary and proportionate measure in a democratic society to safeguard" inter alia: national security; defence; public security and interest; law enforcement procedures; public interest (economic, financial, "monetary, budgetary and taxation matters, public health and social security") of the European Union or of a Member State; to protect "judicial independence and judicial proceedings"; and prevent "investigation, detection and prosecution of breaches of ethics for regulated professions".

In other words, albeit subject to unspecified "necessary and proportionate" measures, these effective exemptions from the application of the GDPR to the processing of personal data cover most governmental agencies of the Member States and EU institutions. The exceptions allow for mass surveillance of EU citizens on any one of the enumerated grounds listed above, including public health, whichever way defined. The exemptions probably also cover non-commercial data brokers employed by agencies and institutions in the interests of governments and public institutions to be safeguarded under Article 23.

**The scope of the GDPR**

Article 68(3) of the GDPR provides for the creation of the European Data Protection Board and twenty-seven Supervisory Authorities (one for each Member State) [63], vested with power to oversee compliance with the GDPR [64] by "controllers" and "processors", and impose sanctions on them for administrative offences.

The term "controller" is defined as "the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law" [64].

Under the EU law, "processors" are natural or legal persons, public authorities, agencies or other bodies which process "personal data on behalf of the controller". Thus, defined, processors include not only major data-driven companies such as Google, Facebook, Uber, eBay, Amazon, or Fuzzy Logix, but also many diagnostic service providers and other health-related businesses.

The compliance regime [66] is based on two major planks: a data subject's right to consent to the processing of his or her data [67], and sanctions for breaches of the Regulations. Central to the compliance regime is "consent", which is defined as:

---

13 Thus, replacing the presently binding Directive 95/46/EC.

14 The EU data protection law does not apply to the data of deceased persons.
“any freely given, specific, informed and unambiguous indication of the data subject’s wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her” [68].

Under Article 7, valid consent must be explicitly provided for processing of personal data and for the purposes of doing so. Data controllers must be able to prove a data subject’s consent, and must not process personal data after withdrawal of consent (though processing of data based on consent before its withdrawal would remain lawful). It must "be as easy to withdraw as to give consent" [69].

For example, a person could have a "pedometer" app on their phone, which obtains user consent as part of its initial setup and then reports on the number of steps the user has walked every day. This information may be then aggregated with other users of the app data and correlated with location, employment, time of day and other factors to create analysis for public health prediction (or marketing) purposes. If the user uninstalls the app and withdraws their consent — the GDPR may require16 that their contribution to the aggregate data set be removed and not used in any further analysis. This is not a trivial technical problem, especially where user data has been datafied into forms that do not readily trace back to an identifiable source.

Supervisory authorities can levy administrative fines of "up to 20,000,000 EUR, or in the case of an undertaking, up to 4% of the total world wide annual turnover of the preceding financial year, whichever is higher" for infringement of "the basic principles for processing, including conditions for consent" [70]. However, the requirement of consent as defined in the Regulation is simply unrealistic in a world where it is predicted that, by 2020, there will be at least 35 billion IoT devices [71]. There are already billions of first generation IoT devices deployed in homes, automobiles, and hospitals in the US [58].

In addition to restrictions on obligations and rights under Article 23, the public interest exception to the requirement of consent in Recital 54 of the GDPR provides that the "processing of special categories of personal data may be necessary for reasons of public interest in the areas of public health without consent of the data subject". The statutory definition of "public health" [72] encompasses "all elements related to health", namely "health status, including morbidity and disability, the determinants having an effect on that health status, health care needs, resources allocated to health care, the provision of, and universal access to, health care as well as health care expenditure and financing, and the causes of mortality". Member States’ electronic health record systems may fall within this exception, and if so would not be subject to data privacy regulation under the GDPR.

The concept and terminology of "privacy by design and by default" is incorporated into Article 25(1) of the GDPR, which imposes a duty on controllers to implement "in an effective manner ... appropriate technical and organisational measures, such as pseudonymisation [2,73], ... data minimisation, ... [and] necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects". The "privacy by design" regulations target the first and second bottom layers of the inverted triangle.

Article 25(2) requires controllers who fall within the third layer of the triangle (and are not exempted by virtue of Article 23) to "implement appropriate technical and organisational" minimisation measures:

"for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility. In particular, such measures shall ensure that by default personal data are not made accessible without the individual's intervention to an indefinite number of natural persons."

While the sentiment of Article 25 is laudable, its mixture of specific and general prescriptions brings to mind the proverbial King Canute's attempt to stop the tide at the sea shore. The "one size fits all" mandatory provisions are financially and technologically onerous, and in relation to small or even medium controllers, may verge on being oppressive [74]. Most probably, multi-national companies like Google, Facebook or Amazon will have the economic, engineering and legal resources to become compliant; whether small and medium-sized healthcare clinics, or start-up medical apps designers will not have funds to reengineer the architecture of their personal data-processing operations is another matter.

Other relevant breaches attracting the Article 83(5) fines include the infringement by controllers of the data subject's "right not to be subject to a decision based solely on automated processing, including profiling [75], which produces legal effects concerning him or her or similarly significantly affects him or her" (pursuant to Article 22)17. Health profiling is precisely the aim of in-database analytic processes of Blue Cross Blue Shield of Tennessee and Crossix Solutions; hence it would be helpful if the US were to adopt a similar prohibition.

However, ban on "transfers of personal data to a recipient in a third country or an international organisation" in Article 83(5)19 is more problematic. Some of the most advanced medical technologies involve automatic

---

16 This depends on the definition of "processing of personal data". If collecting is processing, then the prior data does not have to be deleted. If analysis is processing, then previously-collected data that has not yet been analysed (aka processed) would be subject to removal on withdrawal of consent.

17 Recital 54 further states that "Such processing of data concerning health for reasons of public interest should not result in personal data being processed for other purposes by third parties such as employers or insurance and banking companies".

18 Defences include the data subject's "explicit consent"; contractual requirements as between the data subject and a data controller; authorisation by the EU or Member State law.

19 See also penalties for failure by controllers to provide "transparent information, communication and modalities for the exercise of the rights of the data subject" pursuant to Article 12.
processing to enable collection of the patient’s data from medical tests and sensors to evaluate, analyse his/her health condition, and provide patient-specific recommendations, and predict responses to different treatment options. The relevant data is almost certainly stored on a “cloud” server\(^{20}\), probably in a third country. Even if the treating professional does not adopt the recommendations or the treatment options, there is a risk that this practice may attract Article 83(5) of the GDPR.

Despite the earnest and lofty goal of protecting broadly-defined fundamental rights of data subjects, a centralised bureaucratic approach that relies on litigation\(^ {76}\) and administrative prosecutions to enforce its micro-regulations on data controllers and processors, irrespective of their size and industry sector, is unlikely to succeed\(^ {77}\).

The US model

The US has adopted a different, joint public-private model for safeguarding data subjects’ rights to privacy and security. Like the EU model, its primary object is to protect personal information of data subjects. However, it is based on the notion of data privacy as an aspect of cybersecurity to be safeguarded through the development of integrated government — private sector cybersecurity systems that are “usable, affordable, inherently secure, resilient/recoverable, privacy-protecting, functional, and defensible”\(^ {58}\). Guidelines and standards for “privacy and security by design” are developed by federal agencies in cooperation with the private sector\(^ {78}\).

For instance, focusing on the sharp end of the inverted triangle (manufacturers of internet enabled devices and sensors, apps developers, social platforms, and health service providers), the Report on the Big Data Privacy (“Big Data: Seizing Opportunities, Preserving Values”) by the Executive Office of the President of the US in May 2014\(^ {2}\), recommended six, mainly legislative proposals\(^ {79}\). Then in December 2015, the Commission on Enhancing National Cybersecurity (US) produced the “Report on Securing and Growing the Digital Economy” (Digital Economy Report)\(^ {58}\). It re-phrased the principle of data protection by design:

“Security, privacy, and trust must be primary considerations at the outset when new cyber-related technologies and policies are conceived, rather than auxiliary issues to be taken into account after they are developed. Improved privacy and trust, boosted by transparency and accountability, will contribute to the preservation of civil liberties.”

The US approach emphasizes encouragement of designers and IoT producers to embed safeguards in their products that are scalable and tailored to the function, privacy and security risks of a particular device, app, or service\(^ {77}\). For example, different sets of privacy and security standards pertain to a pacemaker and to a consumer-grade fitness tracker\(^ {58}\). Nevertheless, according to the Digital Economy Report:

“Individuals should not have to be concerned about whether their personal information or information about their behaviours will be tracked without their direct involvement and consent. They should be comfortable knowing that the transmission of information to support identification in an online transaction will be minimized and will not include unnecessary data”\(^ {92}\, p. 17\).

The Report recommends adoption by government and the private sector of strong authentication and data minimization technologies for identity management in apps developed through public-private initiatives\(^ {92}\, p. 16\). It also proposes incorporating into the label (or associated literature) of IoT devices and service providers a standard “rating system based on an impartial assessment of a product’s cybersecurity risk”, including their data privacy-protecting capabilities and practices\(^ {92}\, p. 31\). Such a standard rating system is yet to be developed; however, public-private initiatives “to educate consumers on the selection and use of secure, connected IoT devices”\(^ {92}\, p. 31\) could be created almost immediately. This knowledge would allow customers-cum-data subjects to make better informed choices about the benefits and risks of using health-related technology and services.

Thus, the US approach is based less on policing of compliance with formal consent procedures, and more on embedded controls that enable mitigation of risks to patients’ data privacy and security in health information technology products and services. Developers are also encouraged to build into their products already at the design stage transparency about security and privacy performance. The National Institute of Standards and Technology (NIST) “privacy Control Catalog”\(^ {80}\), provides high-level guidance at the system and design level for the implementation of engineering controls to improve privacy. However, these controls are yet to be incorporated into engineering standards; moreover, NIST’s list of privacy principles and guidelines is still being developed.

Since in the IoT world, the first layer of the triangle is the least amenable to supervisory agencies’ oversight, voluntary compliance with principles and standards for “data protection by design” will be of vital importance\(^ {81}\). It is also at this tier that, for example, the minimisation principle is the most appropriate. For, if the health information technology, including mobile medical apps, limits access to and distribution of personal health-related information about us at this layer, the layers above will have fewer sources to mine.

Under the US model, sector-specific legislation, such as the Health Insurance Portability and Accountability Act 1996 (US), known as HIPAA\(^ {82}\), and Health Information Technology for Economic and Clinical Health Act 2009 (US), known as the HITECH Act, and enforceable rules vested in regulatory federal agencies\(^ {83}\) apply to the second layer of the inverted triangle. For instance, the Health and Human Services Department’s final Security Rule (2003) sets “national standards for protecting the confidentiality, integrity, and availability of electronically[ally] protected health information”\(^ {84}\). As in the EU, US agencies such as the Department of Health and Human Services (HHS), Office for Civil Rights

---

\(^{20}\) The term “cloud” refers to using a network of remote servers hosted on the Internet (rather than a hard drive) to store and access data.
(OCR) can impose administrative penalties for violations of the HIPAA Privacy and Security Rules and the HITECH Act privacy and security provisions [85].

Theoretically, data subjects who suffer compensable injury occasioned by improper processing of their health-related personal data also retain the right to sue for damages at common law. However, for this avenue of redress to become practical, the law needs to re-conceptualise fundamental notions of compensable personal damage, the ambit of duty of care and causation in the age of "Big Data" from the perspective of data subjects. Such re-conceptualisation can be best accomplished through legislation enabling data subjects to sue layer two health-related organisations as well as layer three data brokers and other third party analytics firms and processors.

This proposition is not revolutionary. For example, the Fair Credit Reporting Act (FCRA) § 1681 (15 U.S.C.) enforces persons whose privacy interests have been infringed to pursue private right of action to enforce its provisions. Under § 616 and § 616 of the FCRA, successful consumer-plaintiffs can recover actual damages; and "the costs of the action together with reasonable attorney's fees as determined by the court" for "willful or negligent noncompliance" with any requirement imposed under the Act. If the plaintiff establishes willful noncompliance, the court may allow also punitive damages. The United States Court of Appeals for the Third Circuit in In re: Horizon Healthcare Inc. Data Breach Litigation, determined that for the purposes of the FCRA the intangible harm in the form of loss of privacy occasioned by alleged disclosure of personal information creates a de facto injury without the plaintiff having to demonstrate that the information was in fact used improperly. It is arguable that legislation along the lines of the FCRA, adapted to Big Data healthcare environment would enable members of the Blue Cross Blue Shield of Tennessee health insurance fund tagged as potential opioid drug addicts to sue for damages.

At the third layer of the triangle, data brokers have been the subject of Federal Trade Commission (FTC) actions, and the 2014 Report on "Data Brokers: A Call for Transparency and Accountability" (Data Brokers Report). The Data Brokers Report recommended the enactment of specific legislation that would "enable consumers to easily identify which data brokers may have data about them and where they should go to access such information and exercise opt-out rights". Legislative requirements would include:

- the creation of a centralized mechanism, such as an Internet portal, where data brokers can identify themselves, describe their information collection and use practices, and provide links to access tools and opt-outs; i.e., obligations of "data brokers to clearly disclose to consumers (e.g., on their websites) that in addition to the raw data held for each person (e.g., the person's name, address, age, and income range), they also derive from the data certain data elements", including "sensitive information—and inferences about sensitive consumer preferences and characteristics—such as those relating to certain health information". Consumers should be provided access to derived personal information held by data brokers;
- obligations of data brokers "to disclose the names and/or categories of their sources of data, so that consumers are better able to determine if, for example, they need to correct their data with an original public record source";
- obligations of "consumer-facing entities to provide a prominent notice to consumers that they share consumer data with data brokers and provide consumers with choices about the use of their data, such as the ability to opt-out of sharing their information with data brokers";
- protection of sensitive health-related information "by requiring that consumer-facing sources obtain consumers' affirmative express consent before they collect sensitive information".

The above five disclosure and access obligations, as well as the requirement of explicit consent to sensitive data collection are not dissimilar to those imposed on controllers and processors under the GDPR; however, the US model is more focused, better finessed, and more amenable to policing. That said, the US model for the protection of health-related data privacy at each of the layers of the inverted triangle is, with few exceptions, still at the stage of recommendation for legislative action. This means that, apart from litigation at common law, unless provisions of the HIPAA Privacy and Security Rules and the HITECH Act are engaged, data subjects must rely on voluntary adherence by private and public sectors to non-binding principles and rules.

The private sector's adoption of privacy engineering principles varies widely between companies. For example, in Nokia's submission to NIST the company outlines a mature approach using their own Privacy Engineering & Assurance Process (PEAP) which "is a set of proactive engineering activities to identify the privacy impact of a given object, to design controls and mitigation to ensure appropriate Privacy by Design". The aim is to build "privacy into the object as part of its complete product creation life-cycle. The object can be a software or a hardware product, service, process, operation or even a contract or a partnership". The Nokia's process includes a requirement that the implementation (completion and operability) of the Privacy by Design controls be verified and documented as evidence of "regulatory compliance and in the event of a privacy breach". Nokia has not identified any legislative requirement driving their approach, but states that "[t]he adoption of a Privacy Engineering & Assurance discipline is a necessity in the era of the internet and web".

Intel's submission to NIST was less enthusiastic about the development of privacy engineering standards. This submission stated, "[I]ntel appreciates the interest that NIST has taken in privacy engineering. ... The policy that would be the subject of a privacy engineering standard is not yet settled. ... These and other questions about privacy are not yet resolved, and standards development in these areas would be premature." It is very difficult to quantify the number of companies willing to adopt and implement privacy engineering designs. However, as we, the data subjects, become acquainted with both the nature of the exploitative technology and the technological remedies available, companies and businesses may have to adopt in-built effective privacy safeguards under the pressure of regulation and market forces.
Tests of proportionality

Both the prescriptive EU model and the more collaborative, but still inchoate US approach, essentially adopt, to greater or lesser extent, Ann Cavoukian’s foundational “privacy by design” principles either explicitly in the GDPR, or implicitly in the Digital Economy Report, the Data Brokers Report, the HIPAA Privacy and Security Rules, and theHITECH Act. However, neither model provides a set of tests against which these foundational principles can be evaluated. Neither model provides a clear test capable of being consistently applied within and across each layer of the inverted triangle — from designers of devices, apps and digital systems to the remote third parties accessing digital “composites” of our life and health.

The third approach, based broadly on methodology adopted in the Investigative Powers Act 2016 (UK), involves tests of proportionality and necessity that are devised to safeguard privacy and other significant legal interests of data subjects. Tests of proportionality and necessity are capable of being systematically incorporated into the layers of the inverted triangle, and are fully compatible with both the EU and US models. The concept of proportionality in law has a long history [96]. The public law principle of proportionality germane to the present analysis has its roots in the nineteenth-century Prussian administrative law’s endeavour to protect constitutional rights of individuals from encroachment by the government. Tests of proportionality were used by Prussian courts as an instrument to examine “the legitimacy of government intervention in economic and social life” [97]. In the context of administrative and constitutional law, these tests of proportionality are as follows:

“Whenever the government infringes upon a constitutionally protected right, the proportionality principle requires that the government show, first, that its objective is legitimate and important; second, that the means chosen were rationally connected to achieve that objective (suitability); third, that no less drastic means were available (necessity); and fourth, that the benefit from realizing the objective exceeds the harm to the right (proportionality in the strict sense)” [96].

This articulation of proportionality as four tests is simple and clear, “standard-based rather than categorical, and it is results-oriented rather than being a formal and conceptual doctrine” [98].

Nowadays, the public law doctrine of proportionality has been adopted in many countries [96], though the US constitutional jurisprudence tends to espouse the “law of balancing”, [97,98] a test that balances rights, competing interests; benefits and values rather than determining their proportionality [96,99]. Although the historical, political and theoretical provenance of proportionality and balancing constitutional doctrines is quite different [96], in the non-constitutional context of practically incorporating privacy rights into information systems, these tests can be formulated as the following four rules:

1. On the balance of the specified competing public interests, it is, or it is not, necessary to limit or infringe legal rights of individuals (such as privacy) by legislation, measure, or action in order to achieve a legitimate specific objective.

2. If necessary, the proposed limiting or infringing measures/actions (for example, collection and processing of data) must be rationally connected to (appropriate for) achieving the specific objective.

3. There are, or are not, effective alternative means of achieving the same specific objective without over-limiting or significantly infringing the legal rights of individuals (for example, in the case of privacy, adopting a more narrowly focused collection of data).

4. Deleterious effects of limiting or infringing particular privacy rights of individuals in order to achieve the legitimate specific public interest objective outweigh, or do not outweigh, realisation of this public interest objective.

Under the Investigatory Powers Act 2016 (UK), tests of proportionality and necessity focusing on privacy protections are at the core of the warranty (warrants, authorisations and notices) system that governs “bulk” powers of the national security and law enforcement agencies to intercept communications; interfere with electronic equipment; and access, collect, examine, and manage massive volumes of “bulk” (unfiltered) datasets with “Big Data” characteristics. In order to obtain a warrant, authorisation or notice, the requesting entity has to provide evidence (justify) that the particular measure/action is necessary for fulfilling the listed statutory purpose [100]. The issuing authority must then determine whether the conduct (interception, access, collection, etc) authorised by the warrant, authorisation or notice is “proportionate to what is sought to be achieved by that conduct” [101]. Section 2 of the Investigatory Powers Act 2016 (UK) provides a non-exhaustive list of “general duties in relation to privacy” that those with power to issue the requested instrument must consider when applying the proportionality test. These duties or factors relevantly include:

- data access/collection minimisation “(a) whether what is sought to be achieved by the warrant, authorisation or notice could reasonably be achieved by other less intrusive means”;
- differentiation of protective privacy levels “(b) whether the level of [privacy] protection to be applied in relation to any obtaining of information by virtue of the warrant, authorisation or notice is higher because of the particular sensitivity of that information”;
- balancing the tension between two public interests: “(c) the public interest in the integrity and security of telecommunication systems and postal services” and “any other aspects of the public interest in the protection of privacy.”

Part 3 – proposed framework

While the Investigatory Powers Act 2016 (UK) tests of proportionality, necessity and balancing were developed for government agencies, that is, the fourth layer of the inverted triangle, this approach can be adapted to all healthcare layers within the inverted triangle.
For example, at the first layer, software engineers, computer programmers and apps designers in collaboration with lawyers and healthcare practitioners, can incorporate into systems collecting identifiable personal data precise instructions based on privacy-oriented proportionality and necessity tests, privacy laws, and ethical standards. This can be done by:

- defining the function of the health-related device, app, or service in assisting individual and/or public health;
- assessing the level of security and integrity needed to optimise this function;
- classifying the health-related data to be collected according to its privacy sensitivity level; and delineating appropriate algorithmic parameters for the nature and volume of data to be collected to fulfil the specific purpose, while incorporating input-output authorisation controls according to the privacy sensitivity level.

The actual definition the health-related device, app, or service function might raise an interesting question: does Google qualify as a service that assist public health? The response may well be "yes" — its search engines (algorithms) allow easy access to information (not necessarily accurate) about medical and psychiatric conditions and treatment; they may also provide warnings of potential epidemics.

Nevertheless, if the above measures were implemented at the first layer, there would be less identifiable data health-related available for access and processing at the second layer. These measures can be regulated through the development and use of engineering standards, as is common in non-software engineering disciplines relating to public safety, and increasingly within the software engineering realm.

The second layer consists of clinical and other health-related data in identified form collected, stored and distributed to third parties (medical facilities and public agencies that include but are not limited to healthcare, law enforcement and national security bodies). Algorithms designed for this layer could adapt the four public law proportionality tests focusing on the balance between achieving the specific legitimate objective and limitations imposed on, or exclusion of, privacy rights (as well as on rules for time limits on data storage) [102]. Standards of ethical conduct can also be included in the design, supplemented by voluntary codes and legislation.

Again, if the two lower layers of the inverted triangle were to contain less personal data, then the data brokers of the third layer would have less material to mine. Nevertheless, the raison d'être of data brokers is to profit from algorithmic exploitation of individual data subjects' information without our knowledge and without offering us in return any tangible healthcare benefits. Therefore, data brokers' algorithmic systems, in addition to privacy-oriented proportionality tests for aggregation and processing, will need to be regulated through statutorily entrenched requirements for consent, transparency and accountability.

At the fourth layer, legislative obligations could be imposed on the relevant public and private sector organisations to make the adapted proportionality, necessity and balancing tests part of their information systems.

Finally, the international component of the inverted triangle's fifth layer is yet to be fully developed. Such organisations as Privacy Enforcement Network, which facilitates cross-border cooperation among privacy authorities of its signatory States [103], and regional agreements [104], for example, the Asia-Pacific Economic Cooperation Cross-border Privacy Enforcement Arrangement while useful, are limited in their scope of operation. The ideal would be an overarching and enforceable international treaty on cybersecurity, data protection and privacy. Such a treaty can only be drafted when not just the EU but also the US and other countries are able to articulate a mutually acceptable legal framework.

In the world of Big Data, datafication, the ever more powerful algorithms, and globalised trade in medical devices, apps and services as we as users, customers and data subjects need streamlined privacy protections for health-related information about us that are both reasonably effective and not too onerous to implement. A comprehensive approach that entrenches the "privacy by design" principles adapted through proportionality and necessity tests within algorithmic systems at each layer of the inverted triangle would enhance reliability of privacy protections for the next generation of the Big Data revolution's products.

Disclosure of interest

The authors declare that they have no competing interest.

References

[5] In database processing allows advanced analytics and predictive algorithms (data mining, simulation, forecasting and pattern recognition) "to achieve actionable insights" from numerous data sets without having to import them into a separate analytics environment for processing. In the field of "Preventative Healthcare", the "analysis of 5-10 million lifetime medical records with 2,000 variables" can be performed "in less than 30 seconds." Further Information available at: http://www.fuzzylogix.com; http://www.fuzzylogix.com/wp-content/uploads/FZL-D8lytk.pdf.
Legal protections for personal health information in the age of Big Data


McCarthy M. Apple software will allow researchers to gather health data from iPhones. BMJ 2015;350:h1402.


See https://www. liveathos.com/, see also http:// www.cityzen sciences.fr/en/.

See https://www. liveathos.com/, see also There are many similar apps for all smart phone operating systems on the market. See for example, http://www.my medicalapp.com/; http://www. freehealthtrack.com/; http://www.myhealthdataaap.com/.


Moreover, the fine print of user agreements for apps and extensions often include clauses granting permission to sell data to and access portions of their device by third parties.


The companion instrument, a very lengthy Directive (EU) 2016/680 focuses on enforcement of criminal law.

Regulation (EU) 2016/679, Art 4(15) “personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status”; The Regulation Art 4(13) and Art 4(14) define respectively as “genetic data” as “personal data relating to the inherited or acquired genetic characteristics of a natural person which give unique information about the physiology or the health of that natural person and which result, in particular, from an analysis of a biological sample from the natural person in question”; and “biometric data” as “personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data”.

Regulation (EU) 2016/679, Article 3. The Regulation is also applicable to all foreign companies processing data of EU residents.

Regulation (EU) 2016/679, Article 2(11). Also included is the processing “other than by automated means [i.e., manual] of personal data which form part of a filing system...”. In Article 4(6) “filing system” is defined as “any structured set of personal data which are accessible according to specific criteria, whether centralised, decentralised or dispersed on a functional or geographical basis”. While national electronic health records systems come within definition of a filing system, they are regulated by Member States.


Regulation (EU) 2016/679, Art.2(2)(3). The scope of the GDPR’s oversight does not apply, among others, to the
processing of personal data in the course of an activity which falls outside the scope of the EU law; any activity performed by Member States when carrying out activities in relation to the common foreign and security policy of the EU; and activities by “competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security”; and any processing by the EU itself.


[67] Regulation (EU) 2016/679, Articles 37–39. Controllers and processors (public authorities and r companies processing more than 5 000 data subjects within 12 months) must appoint Data Protection Officers to ensure compliance.

[68] Regulation (EU) 2016/679, Article 4 (11); see also Recitals 31; 42; and 43.


[73] According to Regulation (EU) 2016/679, Art 4 (5): “pseudonymisation means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person”. The problem with such measures is that “de-identification may strip the data of both its usefulness and the ability to ensure its provenance and accountability”. See also: Anderson R. Medical Confidentiality and the Data Protection Regulation. Available at: https://www.cl.cam.ac.uk/~rja14/Papers/Med Confidentiality-and-DPR.pdf; Simon L. Xu W. and Anderson R. Don’t Interrupt Me While I Type: Inferring Text Entered Through Gesture Typing on Android Keyboards. Proceedings on Privacy Enhancing Technologies. 2016;3:136–54.

[74] Regulation (EU) 2016/679, Recital 85 provides that “as soon as the controller becomes aware that a personal data breach has occurred, the controller should notify the personal data breach to the supervisory authority ...not later than 72 hours after having become aware of it, unless the controller is able to demonstrate, in accordance with the accountability principle, that the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons”. See Article 33. Article 34(1) further requires that when “the personal data breach is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall communicate the personal data breach to the data subject without undue delay”.

[75] Regulation (EU) 2016/679, Art 4 (4) defines “profiling” as “any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements”.

[76] Under Article 82 cl 1, “Any person who has suffered material or non-material damage as a result of an infringement of this Regulation shall have the right to receive compensation from the controller or processor for the damage suffered”. Clause 2 imposes on controllers “involved in processing” liability “for the damage caused by processing which infringes this Regulation”. A processor shall be liable for such damage “only where it has not complied with obligations of this Regulation specifically directed to processors or where it has acted outside or contrary to lawful instructions of the controller”.

[77] Regulation (EU) 2016/679, Recital 50 provides that “processing of personal data for purposes other than those for which the personal data were initially collected should be allowed only where the processing is compatible with the purposes for which the personal data were initially collected”.


[79] The recommendations stipulated that: (1) The Department of Commerce should devise draft legislative text for the Consumer Privacy Bill of Rights reflecting risks to privacy presented by Big Data; (2) The Congress should pass National Data Breach Legislation to provide for a single national data breach standard; (3) The Office of Management and Budget should amend the Privacy Act of 1974 to extend Privacy Protections to non-US persons; (4) The federal government should “ensure that privacy regulations protect students against having their data collected in an educational context (schools) being shared or used inappropriately; (5) “The federal government’s lead civil rights and consumer protection agencies should expand their technical expertise to be able to identify practices and outcomes facilitated by big data analytics that have a discriminatory impact on protected classes, and develop a plan for investigating and resolving violations of law”; (6) The Congress should amend the Electronic Communications Privacy Act to ensure that the standard of protection for online, digital content is consistent with that afforded in the physical world. Executive Office of the President, Big Data Privacy Report, Big Data: Seizing Opportunities, Preserving Values, The White House. 2014. p. 60. https://www.whitehouse.gov/sites/default/files/docs/big data privacy report may 1 2014.pdf.


[82] HIPAA's data privacy and security provisions for safeguarding medical information need to be updated. CrossFit Solutions' access to clinical information about patients is based on its patented ‘double-blinded, privacy-safe, distributed data-mining protocol, ensuring that [its] clients have confidence in ... de-identified, HIPAA-compliant approach’. http://crossfit.com/platform.aspx.

[83] For example, the National Institute of Standards and Technology, the Federal Trade Commission; and the Office for Civil Rights of the US Department of Health and Human Services. HIPAA Privacy Rule 2000 45 CFR, Part 160 and Subparts A and E of Part 164, https://www.hhs.gov/hipaa/for-professionals establishes national standards to protect individuals' medical records and other personal health information (it applies "to
health plans, health care clearinghouses, and those health care providers that conduct certain health care transactions electronically).

[84] Details available at: https://www.hhs.gov/hipaa/for-professionals. See also HIPAA Breach Notification Rule, 45 CFR §§ 164.400-414, requires HIPAA covered entities and their business associates to provide notification following a breach of protected health information. https://www.hhs.gov/hipaa/for-professionals/breach-notification/. Similar breach notification provisions are enforced by the Federal Trade Commission under Health Information Technology for Economic and Clinical Health Act 2009 (US), § 13407; known as HITECH Act. The 16 CFR Part 318 Rule requires vendors of personal health records and related entities to notify consumers following a breach involving unsecured health information. In the case of certain breaches involving 500 or more people, the media also has to be notified.

[85] Details available at: http://www.hhs.gov/hipaa/for-professionals/compliance-enforcement/index.html. For example, August 4, 2016 Advocate Health Care Network (Advocate) settlement with the U.S. Department of Health and Human Services, Office for Civil Rights, for multiple potential violations of the HIPAA involving electronic protected health information (demographic information, clinical information, health insurance information, patient names, addresses, credit card numbers and their expiration dates, and dates of birth), included a corrective plan action and payment of $2,140,500 http://www.hhs.gov/hipaa/for-professionals/compliance-enforcement/agreements/sjh).


[87] In re: Horizon Healthcare Inc. Data Breach Litigation, No. 15-2309 (3d Cir 2017) Jordan and Vanaske JJ, with Schwartz J concurring. The case involved theft in 2013 of two laptop computers belonging to the health insurer Horizon Healthcare Services, Inc. The computer databases contained unencrypted personal information about some 839,000 insureds. Horizon offered one year of credit monitoring and identity theft protection services to those affected; however, the plaintiffs alleged that this was an inadequate remedy for their loss of privacy. Available at: http://www2.ca3.uscourts.gov/opinions/152309p.pdf.

[88] In order to sue under Federal statutes, plaintiffs have to establish standing under Article III of the Constitution by demonstrating: “(1) an “injury in fact,” which is neither conjectural nor hypothetical, (2) causation, such that a causal connection between the alleged injury and offensive conduct is established, and (3) redressability, or a likelihood that the injury will be redressed by a favorable decision”. Lujan v. Defenders of Wildlife, 504 U.S. 555, 560–61, 112 S. Ct. 2130, 119 L. Ed. 2d 351 (1992).


[91] See for example, Federal Trade Commission (US) v SiteSearch Corporation, dba LeapJab, et al. No. 2:14-cv-02750 (February 18, 2016), the defendants settled a dispute with the Federal Trade Commission for $5.7 million in collective monetary judgements and were prohibited from further selling or transferring consumer data to third parties or misleading customers about loan application or offer terms. The Federal Trade charged that the data brokers “knowingly sold consumer social security numbers, bank account details, and other information to third parties, who employed this information for illicit purposes”. http://www.whitecase.com/publications/article/ftc-settles-data-brokers-sale-consumer-data-used-illicit-purposes.

[92] Introduced on 10 January 2017, “A bill to ensure appropriate spectrum planning and interagency coordination to support the Internet of Things”. S.88 115th Congress (2017–2018) proposes, inter alia, establishment of a steering committee within the Department of Commerce to advise on “policies or programs that would (i) promote or are related to the privacy of individuals who use or are affected by the Internet of Things; (ii) may enhance the security of the Internet of Things; (iii) may protect users of the Internet of Things; and (iv) may encourage coordination among Federal agencies with jurisdiction over the Internet of Things”, https://www.congress.gov/bill/115th-congress/senate-bill/88.


[100] The national security: preventing or detecting serious crime; the “economic well-being of the United Kingdom so far as those interests are also relevant to the interests of national security” giving effect to mutual assistance instruments and agreements. For example, Investigatory Powers Act 2016 (UK), s 15; s 19; s 21; s 22; s 40; s 63; s 102; s 106; s 117; s 132; s 138; s 158; s 163; s 178; s 204; s 205; s 223; and s 229. Under s 61(7)(e) authorisation to obtain communication s data for “for the purpose of protecting public health” and under s 63(7)(g) “for the purpose of preventing death or injury or any damage to a person’s physical or mental health, or of mitigating any injury or damage to a person’s physical or
mental health**. The legislation also establishes an oversight scheme involving the Investigatory Powers Commissioner and Judicial Commissioners.

[101] Investigatory Powers Act 2016 (UK), s 23(1)(b); s 140 (1)(b); 179 (1)(b); s 208 (1)(b); and s 253 (1)(b).


[103] Details available at: https://www.privacyenforcement.net/.