



Mini Review

Informed Consent in Biomedical Research

Fida K. Dankar^{a,*}, Marton Gergely^a, Samar K. Dankar^b^a College of IT, UAEU, Al Ain, P.O.Box 15551, United Arab Emirates^b Independent Scientist, Ottawa, Canada

ARTICLE INFO

Article history:

Received 19 December 2018
Received in revised form 19 March 2019
Accepted 21 March 2019
Available online 25 March 2019

Keywords:

Informed consent
Privacy
Biomedical research
Biomedical data
Medical care

ABSTRACT

Informed consent is the result of tumultuous events in both the clinical and research arenas over the last 100 years. Throughout this time, the notion of informed consent has shifted tremendously, both due to advances in medicine, as well as the type of data being gathered. As such, informed consent has misaligned with the goals of medical research. It is becoming more and more vital to address this chasm, and begin building new frameworks to link this disconnect. Thus, we address three goals in this paper. First, we discuss the history of informed consent and unify the varying definitions of the term. Second, we evaluate the current research on the topic, classify them into themes, and attend to the problems therein. Lastly, we employ these themes of informed consent research mentioned previously to provide guidance and insight for future research in the arena.

© 2019 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Contents

1. Introduction	463
2. The Evolution of Informed Consent	464
2.1. Informed Consent in Medical Care	464
2.2. Informed Consent in Human-Subject Research	464
2.3. The Definition of Informed Consent	465
3. Current Practice	467
3.1. Current Data-Intensive Research Context	467
3.2. Big Data: Sharing and Risks	467
3.3. Challenges to Anonymization	468
3.4. Challenges to Adapting Consent	468
3.5. New Approaches to Consent	469
3.6. Alternatives to Consent	470
3.6.1. Ownership-Based Governance	471
3.6.2. Solidarity/Community-Based Governance	471
3.6.3. Technical Methods	471
4. Summary and Outlook	472
Acknowledgment	472
References	472

1. Introduction

The notion of informed consent is the outcome of painful abuses in the clinical and research arenas spanning nearly a century. Over time, the underpinnings of the idea of informed consent have shifted, both due to advances in medicine, as well as the type of data being gathered. Thus, informed consent has misaligned with the goals of medical

* Corresponding author.

E-mail addresses: fida.dankar@uaeu.ac.ae (F.K. Dankar), mgergely@uaeu.ac.ae (M. Gergely), samardankar@gmail.com (S.K. Dankar).

research. As such, it is becoming more and more vital to address this gap, and begin building new frameworks to bridge this rift. Our focus in this paper is to familiarize the reader with the broader history of informed consent, introduce the modern aspects of informed consent within the observational medical research arena (i.e.: research using data), discuss inherent problems therein, and begin offering solutions on how to navigate today's informed consent minefield.

Our current understanding of informed consent is that it is the full disclosure of the nature of the research and the participant's involvement, adequate comprehension on the part of the potential participant, and the participant's voluntary choice to participate [1]. However, originally consent was sought for a single study with a pre-specified timespan and specific purpose [2]. With the rise of big data, and the enormous biomedical data warehouses being built, it becomes more and more difficult to foresee the uses and applications of subject data, therefore compounding the difficulty in attaining informed consent, according to the original definition.

Further challenges have been placed on biomedical research through the nature and types of data being gathered. Some biomedical data cannot be anonymized while still retaining utility. Further, some types of data, such as DNA data cannot ever be fully anonymized. In the face of these challenges, we hear increasing calls from prominent scientists and authoritative figures for compulsory research participation [3,4]. They see that individuals have an obligation to participate in research, as it is a benefit for society as a whole [5]. These advocates are pushing for removing the need for informed consent all-together.

As such, the goals of this paper are threefold. First, we aim to address the history of informed consent and unify the varying definitions of the term. Then, we evaluate the current research on the topic, categorize them into overarching streams, and address the issues and challenges therein. Finally, we utilize the streams of informed consent research mentioned previously to assist and guide future research in the area.

2. The Evolution of Informed Consent

Over the last century, the design of informed consent has arisen from tumultuous events in two distinct arenas: medical care and human-subject research. While these streams can be traced back millennia, to the time of Hippocrates, and to the roots of western medicine (when most information was concealed from patients), it is the last 100 years that have been the most horrendous [6].

2.1. Informed Consent in Medical Care

As mentioned prior, the main focus of this manuscript is discussing informed consent in biomedical research. While not immediately related to research-at-hand, informed consent in medical practice is still a vital subset of the evolution of the modern form of informed consent in medical research. Issues with informed consent in the medical care arena preceded the larger-scale devastating breaches of informed consent in medical research, and began to form the groundwork for many of the laws governing medical data today. As such, the history of informed consent in medical practice must be discussed prior to that of medical human-subject research.

Informed consent in medical care is generally regarded as a result of doctors' paternalistic approach to patient treatment. Numerous well-cited court cases were initiated due to adverse effects of medical procedures that patients either opposed, did not approve, or were not provided with adequate information about. Despite this, medicine has a long-standing tradition of upholding patient interests. As far back as 1665, physicians and surgeons were required to obtain a patient's consent prior to treatment under the Duke of York's laws of 1665, which were the founding laws for New York [7,8]. However, even then, these laws had murky grey areas. It was also stated in the Duke of York's laws of 1665 that as long as a treatment demonstrated no perceived

risk or harm, physicians had the right to act without a patient's consent. What constituted 'harm', was up to the physician's discretion.

The first major hurdle leading to informed consent can be traced back to the 1905 case of *Pratt v. Davis*, where a surgeon removed an unknowing woman's uterus and ovaries, in order to treat her epilepsy [8,9]. The physician, Dr. Edwin Pratt, defended his actions by claiming that a patient implicitly consents to a physician when placing themselves in the care of that physician. Further, the physician then has the right to treat a patient as they best see fit with regards to the patient's well-being. Despite Dr. Pratt's claim, the court ruled in favor of Mrs. Davis, arguing that a "physician or surgeon, however skillful [...] [cannot] violate without permission the bodily integrity of his patient" [9].

Despite this leap in favor of the patient's rights, the fundamental turning point for informed consent did not come about until nearly a decade later, with the 1914 case of *Schloendorff v. Society of New York Hospital* [6,8,10,11]. A woman entered the hospital complaining of stomach discomfort. When physicians discovered a mass during examination, the patient insisted that she did not want the mass or tumor excised. Despite Mrs. Schloendorff's refusal, the surgeons proceeded with the procedure, and performed a hysterectomy. Further, due to complications arising from the surgery, the patient developed gangrene on her left arm, and some of her fingers were amputated. Further, additional surgeries needed to be performed on her arm, due to an embolism resulting from the original surgery. In the verdict for this case, Judge Benjamin Cardozo succinctly gave one of the most-cited statements in American medical law, later becoming one of the pillars of current health legislation: "Every human being of adult years and sound mind has a right to determine what shall be done with his own body [...]" [10,11].

The last of the landmark cases in the area of informed consent was the somewhat more contemporary, 1957 case of *Salgo v. Leland Stanford Jr. University Board of Trustees* [6,8]. It was during this case that the term 'informed consent' first appeared. The patient, Martin Salgo, awoke paralyzed after a routine aortography, which proved permanent. Salgo had never been informed by the physicians and surgeons that such a risk even existed, sued the hospital, and was awarded \$250,000 in damages. This pivotal case helped engrain in modern medical care that the failure to disclose medical risks and alternative treatments are cause for legal action [6].

Over time, it was instances such as these that developed and molded the foundation of modern medical legislation, and the notion of informed consent in the patient-physician relationship.

2.2. Informed Consent in Human-Subject Research

Unfortunately, it was not just isolated court cases that guided the evolution of informed consent, but much larger and more catastrophic events as well, upending not only general ethics, but overall human rights. There were three specific events with tragic outcomes that were commenced under the absence of informed consent and external scrutiny. In conjunction, these led directly to the creation of numerous pieces of legislation, as well as both the Belmont Report and Institutional Review Boards.

The first event follows the Nazi party's torturous medical experimentation throughout World War II [8,11,12]. Herein, Nazi physicians tortured and killed concentration camp victims in the name of scientific research. This 'research' was comprised of such horrific acts as killing Jews for anatomical studies, and utilizing non-consenting prisoners for experiments involving exposure to deadly diseases, poisons, simulated high-altitude, and immersion in freezing water [8]. After the end of the Second World War, in 1946, these heinous acts were tried before a tribunal of three American judges in a trial known as 'The Medical Case'. This led to the accountability of the responsible medical doctors, and the introduction of the Nuremberg Code, wherein it was stated that "the voluntary consent of the human subject is absolutely essential", that research subjects "should be so situated as to able to exercise

free power of choice”, and that they “should have sufficient knowledge and comprehension of the elements of subject matter” [13]. While the Nuremberg Code was incorporated into numerous human rights legislations, it did little to guide the behavior of researchers dealing with human-subjects. However, over time, its indirect influence prompted the medical research community to develop its own guidelines for human-subject research, which later became the Declaration of Helsinki [8]. Furthermore, the Declaration of Helsinki has been continuously updated over time, as needed, with its latest update in 2013 [14]. It now includes provisions on the privacy of identifiable human materials and data, as well as the dissemination of their results.

The second instance in which the lack of external scrutiny led to disastrous results was the Stanford Prison Experiment of the 1971 [12]. Philip Zimbardo, a researcher at Stanford University devised an experiment in which a group of students were to act either as prisoners or guards in a makeshift penitentiary setting [15]. The purpose of the experiment was to study “the development of norms and the effects of roles, labels, and social expectations in a simulated prison environment” [15]. After only seven days, the experiment was halted. Following the study, the subjects experienced enduring psychological trauma, and even the researchers admitted to the prolonged psychological effects of the experiment [15].

Furthering the limelight of unethical human-subject research, details of the Tuskegee Syphilis Experiment emerged immediately after the Stanford Prison Experiment [8,12]. In this experiment by the United States Health Service, the effects of untreated syphilis was studied using human subjects over a 40 year span (between 1931 and 1972), despite the subjects being told the experiment would last only a few months [16]. Six hundred impoverished African American land workers were offered the ability to participate in an experiment. Out of this group 399 had syphilis prior to the start of the experiment, and 201 did not [16]. The subjects were not informed that they had the disease, and regardless of having a known cure (penicillin), were not offered any remedies. Instead, infected subjects were told that they were being tested and treated for ‘bad blood’, a term commonly used at the time to refer to a group of illnesses. The subjects were notified that they would receive food, medicine and burial insurance in case of death, in exchange for their participation. However, no medicine was actually given to alleviate the disease [16]. This monumental violation of human rights in research, along with the Stanford Prison Experiment, established the groundwork for the Federal Policy for the Protection of Human Subjects, also known as the Common Rule, which required human-subjects to be overseen by certain governing bodies (which later became Institutional Review Boards or Ethical Review Boards) [12].

In addition, the cumulation of these three cases led to the creation of the 1979 Belmont Report, which detailed the principles through which oversight processes for research on human subjects would be developed [8,12]. Further, the Belmont Report also detailed the basic ethical principles in human subject research: respect for person, beneficence, and justice. ‘Respect for person’ entails respecting the autonomy of all individuals involved in the human subject study, and that people with diminished autonomy (vulnerable individuals, such as children) need to receive additional protection. The purpose of this tenet is to protect subjects from exploitation. They are to be under no obligation to participate in an experiment, and are free to leave at any time. This principle led to the creation of ‘informed consent’ clauses, in which experimenters are required to disclose the purpose, potential risks, benefits, and any alternatives to the study. The second principle of ‘beneficence’ implies that all human subjects are to be treated ethically, and their decisions respected. This follows the Hippocratic creed under which the risk to subjects is to be minimized, while the benefits to them maximized. Subjects are to be protected, and kept out of harm’s way. Lastly, ‘justice’ details that there must also be a fair distribution of the benefit among participants. It is unethical, for example, to provide a cure to a disease for one test group of individuals, while another test group

with the same disease does not receive any help. The Belmont report was a landmark report and subsequently acted as the cornerstone of institutional research policies.

In sum, due to the atrocities within the human-subject research arena, a number of legislations were created including The Nuremberg Code, The Declaration of Helsinki, the Common Rule, as well as the creation of the Belmont Report. Each of these served as a milestone in the creation of the notion of informed consent. Further, while the studies mentioned above were interventional (where treatments were intentionally withheld from subjects either psychologically or medically), not all medical research falls into this category. Much of the medical research today relating to informed consent and data privacy issues, is observational, with minimal bodily risk to the subjects. Naturally, these two types of research pose differing types and levels of potential harm to subjects, and as such, some legislations have begun to address them separately due to this inherent difference (e.g.: Europe’s new General Data Protection Regulation (GDPR) [17]). The data privacy risks will be further developed in Section 3, after discussing the components of informed consent.

2.3. The Definition of Informed Consent

Over time, numerous definitions of informed consent have arisen. This has made it even more difficult to obtain a subject’s consent within the ever-evolving and continually changing data landscape. As such, in this section, these differing definitions are examined, and the definition used throughout the remainder of the article is rationalized. It is important to note, as historically the term ‘informed consent’ is rather recent, published definitions and explanations of the concept are offered from numerous court cases, and legislations, even if the verbatim words ‘informed consent’ were not used. Instead the focus of the below table hinges on the concept and evolution of the meaning behind informed consent.

As can be seen in Table 1, evolution of the term ‘informed consent’ has shifted drastically over time from being purely medical practice-based, to being research-centric and involving all aspects of human-subject research. Throughout the definitions, three constructs underpin the notion of ‘informed consent’ (Fig. 1):

- Study information
- Subject’s comprehension and understanding
- Voluntary participation

First, the most recent pillar of informed consent, states that it is crucial to disclose all information about a study to the participants [13,17–22]. Further, all risks, however trivial, need to be disclosed to subjects, regardless of the effect they may have on said subject’s willingness to participate in the study [18,19,22]. These items usually include items such as research procedures, the purpose of the research, risks and anticipated benefits, any available alternative procedures, and/or the opportunity for the subject to ask any questions [20].

The second construct of ‘comprehension’ is cited by every statement mentioned in Table 1 [7,9,10,13,17–22]. This subset of informed consent evaluates the mental capacity of subjects, and their ability to fully cognize the information given to them by the researchers, as well as understand the risks and benefits of their participation. Of course, comprehension and information are in some ways interlocked, as in some cases comprehension measures how well an individual has the capacity to understand the information that is given to them. Still, during the evolution of informed consent, it was not necessary to disclose all information to subjects. As such, some sources only mention comprehension without the use of information, implying that individuals need to be coherent, and of sound mind (in general), but without the need for physicians to disclose all relevant information to them [5,7,8].

Table 1
The definitions of informed consent.

Source	Definition
1 The Duke of York's Laws (1665) [7]	That no Person or Persons whatsoever, Employed [...] for preservation of Life or health [...] Exercise any force violence or Cruelty upon, or to the Bodies of any whether Young or old; without [...] Consent of the patient or patients if they be Mentis Compotes: much less [...] Exercise any violence upon or toward the body of young or old one or other, to the prejudice or hazard of the Life or Limb of man, woman, or child.
2 Pratt v. Davis (1905) [9]	[...] under a free government at least, the free citizen's first and greatest right, which underlies all others—the right to the inviolability of his person, in other words, his right to himself—is the subject of universal acquiescence, and this right necessarily forbids a physician or surgeon, however skillful or eminent, who has been asked to examine, diagnose, advise and prescribe (which are at least necessary first steps in treatment and care), to violate without permission the bodily integrity of his patient by a major or capital operation, placing him under anaesthetics for that purpose, and operating on him without his consent or knowledge.
3 Schloendorff v. Soc'y of N.Y. Hospital (1914) [10]	Every human being of adult years and sound mind has a right to determine what shall be done with his own body; and a surgeon who performs an operation without his patient's consent, commits an assault, for which he is liable in damages.
4 The Nuremberg Code (1949) [13]	The voluntary consent of the human subject is absolutely essential. This means that the person involved should have legal capacity to give consent; should be so situated as to be able to exercise free power of choice, without the intervention of any element of force, fraud, deceit, duress, over-reaching, or other ulterior form of constraint or coercion; and should have sufficient knowledge and comprehension of the elements of the subject matter involved, as to enable him to make an understanding and enlightened decision. This latter element requires that, before the acceptance of an affirmative decision by the experimental subject, there should be made known to him the nature, duration, and purpose of the experiment; the method and means by which it is to be conducted; all inconveniences and hazards reasonably to be expected; and the effects upon his health or person, which may possibly come from his participation in the experiment.
5 Salgo v. Leland Standford Jr. University Board of Trustees (1957) [18]	[...] the duty of a physician to disclose to the patient 'all the facts which mutually affect his rights and interests and of the surgical risk, hazard and danger, if any.' A physician violates his duty to his patient and subjects himself to liability if he withholds any facts which are necessary to form the basis of an intelligent consent by the patient to the proposed treatment. Likewise the physician may not minimize the known dangers of a procedure or operation in order to induce his patient's consent.
6 Declaration of Helsinki (1964) [19]	In any research on human beings, each potential subject must be adequately informed of the aims, methods, sources of

Table 1 (continued)

Source	Definition
7 Belmont Report (1979) [20]	funding, any possible conflicts of interest, institutional affiliations of the researcher, the anticipated benefits and potential risks of the study and the discomfort it may entail. The subject should be informed of the right to abstain from participation in the study or to withdraw consent to participate at any time without reprisal. After ensuring that the subject has understood the information, the physician should then obtain the subject's freely-given informed consent, preferably in writing. If the consent cannot be obtained in writing, the non-written consent must be formally documented and witnessed. 1.) Information: Most codes of research establish specific items for disclosure intended to assure that subjects are given sufficient information. These items generally include: the research procedure, their purposes, risks and anticipated benefits, alternative procedures (where therapy is involved), and a statement offering the subject the opportunity to ask questions and to withdraw at any time from the research. 2.) Comprehension: The manner and context in which information is conveyed is as important as the information itself. For example, presenting information in a disorganized and rapid fashion, allowing too little time for consideration or curtailing opportunities for questioning, all may adversely affect a subject's ability to make an informed choice. 3.) Voluntariness: An agreement to participate in research constitutes a valid consent only if voluntarily given. This element of informed consent requires conditions free of coercion and undue influence. Coercion occurs when an overt threat of harm is intentionally presented by one person to another in order to obtain compliance. Undue influence, by contrast, occurs through an offer of an excessive, unwarranted, inappropriate or improper reward or other overture in order to obtain compliance.
8 Common Rule (1991) [21]	[...] no investigator may involve a human being as a subject in research covered by this policy unless the investigator has obtained the legally effective informed consent of the subject or the subject's legally authorized representative. An investigator shall seek such consent only under circumstances that provide the prospective subject or the representative sufficient opportunity to consider whether or not to participate and that minimize the possibility of coercion or undue influence. The information that is given to the subject or the representative shall be in language understandable to the subject or the representative. No informed consent, whether oral or written, may include any exculpatory language through which the subject or the representative is made to waive or appear to waive any of the subject's legal rights, or releases or appears to release the investigator, the sponsor, the institution or its agents from liability for negligence.
9 EU Directive 2001/20/EC (2001) [22]	Informed Consent is the decision, which must be written, dated and signed, [...] taken freely after being duly informed of

Table 1 (continued)

Source	Definition
10 General Data Protection Regulation (2018) [17]	<p>its nature, significance, implications and risks and appropriately documented, by any person capable of giving consent or, where the person is not capable of giving consent, by his or her legal representative; if the person concerned is unable to write, oral consent in the presence of at least one witness may be given in exceptional cases, as provided for in national legislation.</p> <p>1.) Where processing is based on consent, the controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data.</p> <p>2.) If the data subject's consent is given in the context of a written declaration which also concerns other matters, the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language. Any part of such a declaration which constitutes an infringement of this Regulation shall not be binding.</p> <p>3.) The data subject shall have the right to withdraw his or her consent at any time. The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal. Prior to giving consent, the data subject shall be informed thereof. It shall be as easy to withdraw as to give consent.</p> <p>4.) When assessing whether consent is freely given, utmost account shall be taken of whether, inter alia, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract.</p>

Thirdly, voluntariness, has also been mentioned in all 10 publications listed in Table 1 [7,9,10,13,17–22]. This construct emphasizes the necessity of a subjects' consent to be fully voluntary, with no caveats, coercion or influence from the investigator. Further, voluntariness has sometimes included not just the act of stepping into a research or medical study environment, but also the act of withdrawal from it. The notion of voluntary withdrawal from studies and research is rather recent. Until now, it has only been mentioned in the Belmont Report, the Common Rule, the Declaration of Helsinki, and the 2018 General Data Protection Regulation [17,18,20]. As such, it could be argued that other sources only explicitly offer partial voluntariness.

As such, we have combined these three elements to form the definition of 'informed consent' used herein: Informed consent is the full disclosure of the nature of the research and the participant's involvement, adequate comprehension on the part of the potential participant, and the participant's voluntary choice to participate [1].

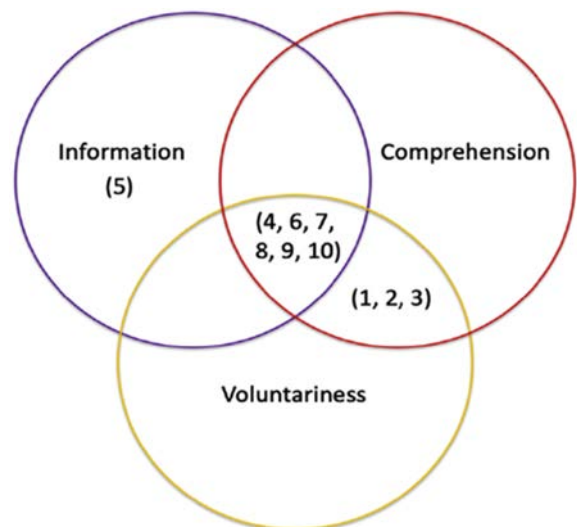
In the next sections, the current practices in informed consent are analyzed, and the possible future directions in the area discussed.

3. Current Practice

3.1. Current Data-Intensive Research Context

Big data is triggering a revolution in Biomedical research and health care. This transformation is fueled by the advancement in digital technologies that can track individual's behavior to a higher degree, and by the technical capacity to store, link, and analyze data [23].

In our context, "big" refers to the volume of the data, its diversity (originating from multiple sources) and/or complexity (the

Fig. 1. The Three Pillars of Informed Consent¹.

computational effort necessary for analysis) [24]. The big data transformation can be witnessed in the biomedical research field which, increasingly requires the continuous collection and linking of a wide range of information types from various observational and instrumental sources such as health data, research data, and data from health devices/services.

The variety in the collected data types along with the advancement in digital technologies are paving the way to novel approaches in biomedical research. Precision medicine is one such data-driven approach that promises to advance diagnosis, prevention and treatment of diseases. The personal Genome Project [25], the 1000 Genome Project [26], Genomics England [27], the Danish National Centre for Register-based research [28], and All of US [29] programs are just a few recent examples of active large-scale projects generating enormous databases that link complex biomedical information, promote sharing and re-use of data, and support continuous data updates.

The collection and management of such a complex bouquet of heterogeneous data can only be justified if the data is kept for long (indefinite) periods, used in multiple studies, for different research purposes, and if it is shared widely with other investigators. This however raises many privacy and ethical issues.

3.2. Big Data: Sharing and Risks

The growing amount of data collected on individuals paints a precise picture about them even when direct identifiers are removed. The indefinite storage, and unlimited re-use of the data aggravates and extends the problem. Moreover, the advances in genomic sequencing and the ability to incorporate Genomic data with clinical data complicates the problem even further and raises new and challenging privacy problems. Genomic data is highly distinguishable (there is confirmation that a sequence of 30 to 80 single-nucleotide polymorphism (SNPs) could uniquely identify an individual [30]), and very stable [31] (genomic data undergoes little changes over the lifetime of an individual, and thus has an eternal value). It provides sensitive information about genetic conditions and predispositions to certain diseases such as cancer, Alzheimer's, and schizophrenia. Moreover, it does not only provide information about the sequenced individuals, but also about their ancestors, and offspring (the authors in [32] describe privacy risks to the family members of individuals who shared their genetic data for medical research). Such personal information can be very damaging to the

¹ Note: The numbering of the articles inside the Figure refer to the numbering structure of the sources in Table 1, as opposed to in the references section.

individuals, and it can be used against them to limit insurance coverage, to guide employment decisions, or to apply social stigma.

The economic harm that results from insurance and employment discrimination is the highest concern cited by 52% of study participants [33]. A 2005 report by Greenhouse and Barbaro alleges that major employers (including Wal-Mart) base some of their hiring decisions on the applicants' health records [34]. Another 1996 survey of people at risk of developing a genetic condition found “more than 200 cases of genetic discrimination among the 917 people who responded” (these cases involved either insurance and/or employer discrimination.) [35]. Another survey of care givers and patients identified “550 people who had been denied employment or insurance based on their genetic predisposition to illness” and that “misuse of genetic information could have intergenerational effects that are far broader than any individual incident of misuse” [35].

While the above examples act as impediment for biomedical data sharing, the demands for data sharing are mounting, and not solely for research facilitation. Scientists are increasingly required to share their data to allow for study replication and verification, and to abide by the rules of the publication medium [36]. Certain federal agencies in the US (such as the National Institute of Health (NIH)) mandate the sharing of data generated or used under federal funding.

Legally, data sharing requires the consent of individuals to the concerned study, or the anonymization of the data. This is conventionally known as the ‘consent or anonymize paradigm’. In fact, in both the United States and Europe, biomedical data can be shared and re-used without seeking additional consent if the data is anonymized.

3.3. Challenges to Anonymization

Anonymization is a data protection technique that is generally used to avoid seeking consent [23]. The Common Rule, for example, allows the use of data and samples “without the knowledge and consent of the individual provided they are de-identified” [21]. The European Union's data protection regime states that the removal or perturbation of information can transform personal data into non-personal data [37]. Similarly, GDPR states that “the principles of data protection should [...] not apply to anonymous information” [17].

Anonymization is legally defined as a technique to prevent identification taking into account all the “means reasonably likely to be used to identify a natural person” [17,37]. It is a process of de-identification that produces data that cannot be linked back to its originators. However, the effectiveness of anonymization has been questioned in the past 15 years as high-profile cases of re-identification occurred when clinical data are re-used. For example, in 2006, America Online (AOL) published a sample of its search queries after replacing screen names with random numbers. The New York Times were able to re-identify and reveal one of the users. Similar and more recent anonymization failures occurred to data released by Netflix [38], the New York Taxi and Limousine Commission [39], and more [40–42]. In fact, there is a consensus among privacy experts that no form of de-identification can guarantee the above notion of anonymity.

Re-identification of anonymous subjects usually occurs by combining various publicly available information with the de-identified data. The type and complexity of biomedical data being collected, along with the abundance of publicly available data sources, contribute to the inadequacy of anonymization even further. This is particularly true when huge amounts of data are collected for individuals over extended periods of time (electronic medical records-EMR data), and even more so, when genomic data is involved.

Multiple independent and recent genomic studies demonstrated the possibility to breach the privacy of participants from genomic datasets that were thought to be anonymized. Gymrek et al. were able to infer the identity of 50 anonymous male subjects using their sequenced Y-chromosome, as part of the 1000 Genome Project [43]. The researchers exposed the identities of these anonymized research participants, as

well as their relatives using public genealogy databases. In other recent studies, Sweeny et al. identified participants in the Personal Genome Project using public demographic data [44], Malin et al. used public information, such as death notices, to recover the identities of family members of participants in the 1000 Genome Project [45], and Homer et al. demonstrated the ability to identify the participants of a genome-wide association study using data from the database of genotypes and phenotypes, dbGaP. In the last example, The authors proved that allele frequencies of a subject can be used to determine whether they participated in a study with high accuracy “even if the study reported only summary statistics on hundreds or thousands of participants” [44,46]. These, and similar studies [47,48], had great impact on the research community due to the increasing sensitivity of the collected biomedical data and its mounting potential for harm.

Although anonymization does not fully protect the data (in that re-identification risk remains), it necessitates extensive stripping of data sets and excludes data updates and linking (once the data is anonymized, there is no possibility to update it with new information nor link it with another dataset), which are essential activities in modern biomedical research. A solution has been to assign a unique random identifier to every subject using a process known as pseudonymization. However, pseudonymization does not preclude the application of data protection law [17].

Moreover, while inadequate in protecting data, anonymization gives legal immunity to data holders and releases them from the obligations of further data protection [23]. Consent on the other hand, is accepted globally as the highest standard of legitimacy, and as most likely to promote and foment trust with participants and communities [49]. In the following section, the challenges in adapting consent to the current data-intensive age are discussed.

3.4. Challenges to Adapting Consent

The consent mechanism should provide data subjects with control over who can access their stored data, for what purposes, and for how long. The challenges to adapting existing consent processes to modern biomedical data are multi-dimensional. Current consents (i) lack the dynamicity required for modern data, (ii) they are not well suited for the education and comprehension components, (iii) they are unable to deal with consent revocation, and (iv) they do not have mechanisms for the return of research results to participants in the context of translational genomics.

- i. Traditional consent was designed to deal with a single study, having a specific purpose, and a pre-defined timespan. The process is mostly paper-based, and is typically logged at the start of the study, resulting in a static process that locks consent information to that single time point [50], and requires all future data usages to be specified at the time of the initial consent. With the rise of big data, data sharing, data re-purposing, and data aggregation into enormous data warehouses, it has become unfeasible to foresee all future uses and applications of the data at the initial time of collection. Pre-defining an array of future uses on the dataset limits creativity, and can act against the incremental nature of research.
- ii. The ‘information’ and ‘comprehension’ constructs of informed consent (as mentioned in Fig. 1) impose the inclusion of a well-designed education and assessment component. These components ensure that all information related to the data usage and data protection are conveyed and understood by the participants. These notions are challenged with current big data for several reasons:
 - a. Our understanding of the data, and what it reveals about the participants is progressing, thus it is not possible to quantify the amount and sensitivity of personal information that can be derived from the collected data.
 - b. Big data is often used to reveal unpredictable connections between data points, as such, it presents greater uncertainty regarding

- future research. An example is provided by Mittelstadt regarding the secondary effect of medications “which can be identified by comparing data not only from multiple clinical trials, but informed sources as well, such as incidental self-reporting via social media and search engine queries”, such connections cannot be “accurately predicted prior to carrying out research” which means that consent cannot be informed [24].
- c. Studies show that participants have discrepancies in their understanding of consent information [51,52], yet consent forms continue to increase in length and complexity and are likely to incorporate difficult legal language [53,54]. In many instances, participants are provided with these lengthy documents just to alleviate any legal responsibility, which does not serve the comprehension construct. With the uncertainty surrounding future research, the long life of modern data, and the obligation to provide participants with information about how their data will be protected [24], it is essential for modern biomedical data warehouses to include a well-planned and comprehensive education program that recognizes that (i) participants (and humans) can only absorb a limited amount of information at any one time, and that (ii) re-contacting participants to provide additional education materials (or to obtain appropriate consents for each new study) is arduous, time-consuming, and expensive. Moreover, it can have a negative impact on the participants, and on the research endeavor.
 - iii. Consent revocation is a necessary motivator for research participation, and an essential part of the ‘voluntariness’ construct. The authors in [55,56] argue that as consent grants individuals the right to decide when and how to use their data, it should logically give them the right to change or withdraw their consent. Traditionally, withdrawing from a study an individual consented to was a straightforward process. Conversely, as current data is re-used and shared with multiple research organizations over indefinite periods of time, it is complicating the issue of revocation significantly. A good tracking mechanism for data-sharing should allow consent revocation provided proper agreements are made prior to sharing the data, and proper mechanisms are used to ensure compliance. However, it is important to note that data withdrawal could severely bias study results, thus compromising the study and undermining the investigators’ effort and time and wasting resources. Current best practices recommend that any samples collected from the individual wishing to withdraw from the study be discarded, and that medical data no longer be used. However, shared samples and data do not necessarily need to be revoked [57]. GDPR for example gives data subjects the right to withdraw consent at any time and states that “it shall be as easy to withdraw consent as to give it” (the previous EU Directive did not address consent withdrawal specifically), however, withdrawal only applies to future data processing and not to data already in process. Yet, it is not clear how to deal with the data of dead participants and whether relatives or offspring can exercise the right to revocation.
 - iv. In this highly dynamic context, the lines between clinical and research practice are fading. Research in translational genomics is increasingly calling for the return of individual results back to participants and their physicians, thus challenging the traditional approach to consent even further. The recommendations governing the return of individual results are usually aligned with returning ‘clinically actionable’ results. Such recommendations are often criticized for excluding the community from the discussion [58]. Multiple studies have deduced that the majority of participants would like to learn more about their genetic results (than is recommended). Further, they would like to decide what results to be returned [59]. This complicates the ‘information’ and ‘comprehension’ constructs of the informed consent once again, and necessitates a re-design of a consent process to capture the informed (and fluctuating) choices of participants with regards to returning their

interpreted data, and to educate participants on these very difficult and highly sensitive issues.

In conclusion, informed consent is evolving from a single contract for a specific experiment, with well-defined purpose, and a limited time span, into a dynamic, long term contract, that requires continuous subject involvement and education. This evolution, and the challenges it puts forward, invokes varying approaches to consent from the research community. We examine these approaches next.

3.5. New Approaches to Consent

The challenges to consent adaptation created a tension between the proponents of open data, and those who advocate for informed consent. Proponents of open data increasingly view consent as a barrier to creative and timely biomedical research [60]. They advocate for the elimination of consent altogether, and in some cases, they go as far as advocating for a ‘duty to participate’ [5,24]. Proponents of informed consent are calling for the preservation of consents’ core values (information, comprehension, and voluntariness), and they suggest using modern IT to maintain active and continuous engagement with participants [61]. As such, various consent mechanisms have been developed, or are under development, (limited by the governing privacy legislations) that span between these two extremes.

The majority of modern data warehouses practice global types of consent such as ‘broad’, ‘blanket’, and ‘tiered’ consent [62]. Blanket consent authorize all future and unforeseeable data uses with no limitation [63]. That is, participants consent to the broad usage of their data, without any specific information about any future research, nor any possibility for further oversight. Broad consent also authorizes all future and unforeseeable data but within broadly specified research areas, and gives subjects the ability to exercise restrictions in some of these areas [63]. In that sense, both mechanisms are a shift from specific consent, into ‘consent to be governed’ (by the institution holding the data) [23]. The recent revision to the Common Rule promotes the use of broad consent as a mean to facilitate research data collection [63]. Broad and blanket consents have been criticized for limiting the autonomy of participants and for not offering them any meaningful control over their personal data and samples [64,65]. Proponents of broad consent argue for its ethicality if certain restrictions are imposed, for example, if personal information is handled safely, if subjects are given the right of withdrawal and if all studies are overseen by an ethics board.

Tiered consent is a minor upgrade from broad consent. It allows participants to choose general research areas they wish to participate in, and exclude others. Subjects can also request to be re-consented for future uses of their data. While tiered consent offers more autonomy for the subjects, it is still limited by the pre-specified (and general) uses of the data. Similarly, ‘opt-out’ is another consent mechanism that assumes broad consent for research by default, unless individuals explicitly opt-out. Opt-out consent models are not seen as ethically acceptable. As pointed out by Hayden [46], they tend to take advantage of people in vulnerable moments, such as when they are seeking treatment. These models have a common goal of maximizing the availability of research data with limited involvement and control from the subjects. As such, they dilute the information construct of the informed consent, for the advantage of research and science. To compensate for consent relegation, some organizations designate institutional review boards and ethics review committees to provide research oversight [66], by reviewing each data usage/sharing instance to ensure that it is not problematic [67,68]. However, the extent of oversight and control from these boards is subjective, and varies from one implementation to another.

Dynamic consent is another approach to consent that claims to be consistent with the three pillars of informed consent. Dynamic consents are personalized online consent and communication platforms [69], and use modern IT to provide a communication channel between researchers and participants of a project. This allows consent documents

to be tied to events in real-time, as they occur in the data life cycle. Such framework enables alerting individuals of new research opportunities, and allows participants to accept or decline participation in a research project if they match the research's specific profile, moreover, it enables sponsors to conform their protocols to the dynamic privacy laws. Researchers will be able to update or include new consent documents whenever significant feedback from participants is required, and participants will be able to respond to these changes and/or to modify their consent at their convenience. The framework also allows all consent material to be available in one platform, including the return of individual results back to participants. For example, if new knowledge/treatments are available that could have an impact on participants (such as if new information is generated that changes a variant's status from ambiguous to actionable), then additional consent documents can be created to allow participants to decide if they want to receive more information about the treatment, and/or to allow that information to be transmitted to their physicians. Recently, several clinical and population-based genomic research projects are implementing dynamic consent platforms [70–73]. Similar projects are underway in the health care arena that enable patients to view, update and share their health records with health care providers [74,75]. Some commercial genomic sequencing companies, such as 23andme [76], provide a limited form of dynamic consent models through their secure online portal systems. While dynamic consent platforms are fairly new and mostly under development, they are being welcomed as empirical evidence shows that a significant number of participants would like to have substantial control over their data. At the same time, multiple concerns have been raised and need to be considered prior to adoption: (i) Dynamic consents are costly to implement and to maintain, (ii) participants enrolling in studies through online portals may not be diverse in terms of race and education level as shown by recent studies [71], (iii) dynamic consents require excessive time from participants to comprehend and consent for each study, this could lead to information overload and withdrawal, it moreover may lead to excessive self-protection behavior [77], (iv) dynamic consent may be asking participants more than they are able to deliver, as they have the responsibility to decide on complex issues they do not have the time, or capacity, to fully comprehend or assess, and (v) dynamic consent does not solve the issue of consent revocation and its implications. As argued in [78,79] withdrawal is complicated by the fact that participants' samples/data may already have been shared by multiple other research organizations.

The comprehension pillar of informed consent is increasingly challenged by the complexity of biomedical research and by the low rates of health literacy [80]. As argued before, informed consent requires asking some people more than they can deliver, given their time and capacity. As a solution, it has been suggested, wherever permissible by law, to delegate consent to a third party, or to one of pre-defined consent models (in such case, representatives or care givers could help participants to decide). Thus, participants would 'consent to be governed by a third party of choice' rather than consent to detailed studies [81].

Open consent is another form of consent that relies on 'information altruists' to openly share their data for the public good [2]. Many investigators have highlighted that the gathering of increasing amounts of clinical and genetic information cannot be made anonymous and will ultimately reveal subjects' identities in our data driven and wired world. The growing availability of public data on the internet including personal (what people reveal about themselves), and genealogical data makes it extremely difficult to share data in a secure way [82]. Open consent proponents realize that promises of privacy may not be realistic, so they rely on individuals who value public good more than confidentiality to donate their data unconditionally for research. By its nature, open consent commands openness and honesty from investigators as well. Participants need to feel passionate about the ultimate goal of the research and they need to be constantly updated about how their participation is facilitating research, and ultimately improving knowledge in the biomedical field. In addition, investigators have a moral

obligation to discuss clearly (as much as possible) the repercussions of biomedical data sharing. The Personal Genome Project, PGP [25], which practices open consent, actively engages its participants by returning all results, and strives to keep them truly informed along the way. The education component in the PGP open consent is extensive, potential participants undergo an eligibility screening to understand whether their privacy preferences conform with the open consent mechanism. They are given study material that offers an introduction to genomics, and that discusses potential risks of data sharing. Individuals are then tested for comprehension, and are required to score 100% to be eligible for participation. The risks discussed with potential participants in PGP include remote scenarios such as "being framed for criminal activities [because of your genetic makeup], or being cloned" [82]. Participation in this case rests on one's faith in the societal benefits of research and is thus "consenting to uncertainty" [83]. However, only people with good social standing tend to openly offer their data. These are people with good social and economic status, that generally offers them protection against any future discrimination or harm² [2,24,85]. Unfortunately, effective result-driven research mandates the inclusion of diverse subpopulations. Skewed samples are not representative of the underlying community and thus lead to inaccurate results.

With the potential gain from big data and the inadequacy of consent, scientists are growing impatient. After all, novelty and experimentation are at the heart of big data research and it is hard to confine these to the norms of a traditional consent. Thus, there are increasing calls from prominent scientists for obligatory sharing of research data (a step beyond scrapping consent altogether). Proponents of this method believe that laws should be instated to protect participants from discrimination based on their health information. An example of such law is the Genetic Information Non-discrimination Act (GINA) adopted by the US government in 2008. GINA forbids discrimination by insurers or employers on the basis of genetic information. The problem with such regulations is that they are enforced only when discrimination on the basis of genetic information is proven, which necessitates the difficult task of knowing/establishing these malicious intentions. So, until such laws are successful in thwarting these risks, it will be hard for this idea to gain traction.

Table 2 below displays the established, modern consent mechanisms mentioned in this section, and evaluates them in terms of their adherence to the pillars of informed consent (information, comprehension and voluntariness). It also provides examples of real-world research projects within each mechanism. Please note:

- 'Varies' (under the Information heading) indicates that the information given to participants depend on the institution implementing the mechanism. For example, the Vanderbilt Genome-Electronic Records Project (VGER) [66], a subset of the eMERGE network [86], implements the opt-out mechanism, however they have an extensive information program to educate and involve the community about their project.
- 'Partial' withdrawal indicates that participants can easily withdraw from any additional usage of their data, however they cannot withdraw from studies already underway.
- '-' indicates no available information on how withdrawal is implemented.

3.6. Alternatives to Consent

Rapid progress is taking place in privacy research in the biomedical area, driven by the need to protect and benefit from big biomedical

² In a 2015 Aljazeera report, an HIV patient lost his job and his health insurance after his health record went public. This occurred after the lab technician informed his insurance company, which later informed his employer. This blatant breach of privacy is common for vulnerable HIV patients in many parts of the world, as HIV is highly stigmatized [84].

Table 2
Consent mechanisms vs. the three pillars of informed consent.

Consent type	Institution	Information	Comprehension	Voluntariness		Remarks
				Participation	Withdrawal	
Broad	1000 Genome Project [26] Estonian Genome Project [87] H3Africa ^a [88]	Varies	×	√	–	Limited autonomy
Tiered	H3Africa ^a [88]	Varies	×	√	–	Limited autonomy
Opt-out	VGER-eMERGE [66]	Varies	×	×	–	Ethical issues
Open	PGP [82]	√	√	√	×	Study bias
Dynamic	All of Us [29] Faroe Genome Project [89] 23andme [76]	√	√	√	Partial	Costly, Information overload

^a H3Africa, the Human Heredity and Health in Africa initiative performs research into the genetic and environmental basis for human diseases relevant to Africans. It involves several projects, 7 of which use broad consent, and 5 used tiered consent.

data warehouses. Some of these methods provide alternatives to ‘informed consent’ through novel governance mechanisms. These will be presented in the following two sub-sections.

3.6.1. Ownership-Based Governance

Many privacy and ethics activists argue for the right of individuals to own their data, in fact, data ownership is at the heart of the ethical debate behind informed consent. While ownership is often discussed as the right of the subjects to control their data, it can also refer to owning intellectual property produced from using the data. Right of data control is defined as the subjects’ (unlimited) right to track all data usages and restrict undesired uses [90]. Naturally, this includes understanding who holds what data, and for what purpose. Data control is also sometimes interpreted to include the right to data guardianship, in other words, individuals hold on to their data, and decide to grant access (to family, health authorities, or research facilities) as they see fit. However, the numerous challenges in terms of data storage data security, and access control, speak against this scheme [62].

Even without data guardianship, significant technical and practical challenges stand against the realization of this concept in the era of big data. For instance, to have any meaningful control, individuals are required to put significant effort to compile and understand their data, to understand the aims of the proposed research, as well as treatments available for any relevant condition(s). As mentioned in [24], “Big data requires significant computational power, storage, and advanced scientific know-how. [...] Analysis will require discipline-specific skills and knowledge, [and] the amount of time and effort required [...] may easily be overwhelming”.

Other issues that require attention relate to data ownership after the subjects’ death, and responsibilities toward a close family member that requests results (particularly in the case of genomic data) [91].

3.6.2. Solidarity/Community-Based Governance

Recent research proposes to revolutionize the notion informed consent and eliminate some of its downside. The proposals in the literature include: (i) basing research participation on the concept of solidarity, (ii) shifting the notion of consent toward a community-based environment, or (iii) granting systemic governance of data to research boards.

Prainsack et al. presented a shift from the traditional autonomy-based consent governance toward a solidarity-based governance mechanism [92]. Solidarity recognizes people’s willingness to engage in risky activities for the benefit of others. They argue against the current governance framework for being solely founded on the promotion of participants’ autonomy, neglecting the fact that assisting others is one of the main motivations of research participation [93]. Their solidarity-based framework would thus “[Assist] others as its main research goal [...] to improve [the] health of individuals” [92]. The application of this novel concept necessitates major shifts in the research practice. The framework is based on individual willingness to accept certain levels of risk and uncertainty for the benefit of others, rather than the

exclusive focus on risk mitigation. General values, mission, and goals of proposed research (what it hopes to achieve) should be communicated to participants, re-consent should only be sought when such goals change significantly. The mechanism also shifts the emphasis from risk mitigation to harm mitigation, meaning that it would devise strategies for action if an undesirable event actually occurs (instead of focusing all resources on risk prevention). For example, funds can be made available to individuals that actually suffer harm, such as victims of discrimination.

One of the main arguments for a solidarity-based consent mechanism is to give up consent in its current form entirely, as it is believed to be hindering the progress of research and failing to benefit society. However, eliminating the need for informed consent amid the potential risks that big data entails is not something that can be accepted without a public debate [24]. Moreover, the uncertainty over what the data will be able to reveal about the participants in the future (particularly genetic data) renders the argument of ‘accepting costs to assist others’ unfounded, and renders individuals accountable of the heavy burden of this social good [94].

Community-based participatory research (CBPR) is a collaborative research between scientists and the community affected directly by the research (such as a group with a common illness) [95]. CBPR is characterized by shared decision making, shared management and ownership, transfer of knowledge and expertise, shared analysis and interpretation of data, and wide dissemination of findings [82]. While CBPR is developing rapidly, it can be complicated and challenging. It requires significant commitment from community members (particularly for pursuing grants) requiring them to give up some of their existing roles. However, there are growing numbers of funding opportunities dedicated to CBPR, including new National Institute of Health (NIH) sponsored funding, as well as fellowship programs, workshops, and courses [95].

Lastly, the ‘systemic oversight’ model calls for stronger oversight mechanisms that are “aimed at orienting and monitoring [...] research activities” [77]. As such, the oversight mechanism should maintain ethical controls throughout the continuum of data activities (data collection and data use), and direct health-research activities toward socially anticipated topics.

3.6.3. Technical Methods

Technical mechanisms attempt to perform data analysis in a privacy preserving manner without the need for consent. De-identification is the most familiar technical approach to privacy. It is a complex process that consists of generalization techniques [96,97], randomization techniques [98,99], or synthetic data generation techniques [100]. However, as mentioned earlier, de-identification techniques are not effective in the modern data context, they significantly reduce the utility of the data, without necessarily offering the required privacy [42,101].

The inadequacy of current de-identification methods, prompted privacy experts to seek new solutions that protect and benefit from the large biomedical data warehouses being built worldwide. Current

research into technical privacy preserving mechanisms are numerous, among the most promising we cite risk-aware systems, and Secure Multiparty Computations (SMC).

Risk-aware systems attempt to quantify privacy risk, both from the participants' perspective, and the data holder's perspective. They then try to impose mitigation measures to counter the posed risk. As such, the systems evaluate the risk posed by a data access request, and provide it with appropriate protection measures [67,102,103]. In essence, more mitigation measures would be imposed on requests of higher sensitivity (requests for data with higher potential for injury or from requester that are not highly trusted). Such mitigations could manifest as a reduction in the granularity of the data and/or as restrictions on when and how a user can access the data [104,105]. Implementations of risk-based systems are found in [106,107].

SMC approaches are based in cryptography. SMC allows multiple parties (each holding a fraction of the data to be analyzed) to collectively carry out calculations on the overall dataset without any party having to reveal any of their own raw data. SMCs offer an attractive solution particularly when coupled with differential privacy [108], however they are not efficient in every setting due to their communication and computational overhead [109,110] (extensive message passing between the different concerned parties is the main bottleneck of existing secure multi-party computations). Relevant implementations of SMC can be found in [110–117].

The search for technical ways to deal with the privacy problem is becoming increasingly challenging in the big biomedical data context. The technologies mentioned above help moderate some of these difficulties, however they still have to overcome significant challenges before being adopted widely.

4. Summary and Outlook

The principle of informed consent was introduced over a century ago, in order to protect patients from unwanted intrusions into "their bodily integrity" [118], and give them the ability to decide on their treatment options. Later, informed consent was extended to research practice in order to grant individuals the autonomy to decide when, how, and with whom to share their data and samples. However, tension around informed consent has existed since its inception. After the Nuremberg code, physician-investigators were reluctant to share their medical/research decisions with their patients, and worked hard to water down the requirement of informed consent as stated by the Nuremberg Code [8]. This could be witnessed in the Declaration of Helsinki, wherein it was stated that for individuals suffering an illness, informed consent "was only necessary [as long as] it was consistent with patient psychology" [8]. It could also be seen in the provisions of the Common Rule, which allow consent exemptions, under the guidance of IRBs.

With the advances of big biomedical data, along with the challenges they posed to consent, the friction between proponents of open data and proponents of informed consent is escalating.

Proponents of informed consent are calling for a refocus on the pillars of informed consent, in order to avoid a replication of the past. However, one of the main challenges of applying informed consent in big data is the information overload. Significant amount of time and effort is required from participants to attain the required knowledge and background that enables them to understand the totality of the data about them. While there are many willing individuals, the process will be overwhelming and highly stressful for most. A good solution could be in allowing a multi-consent mechanism that enables subjects to choose whether they want to acquire the needed background, to delegate their consent to a trusted and independent third party (consent to delegate data governance to a person of choice), or any other consent mechanism that best suits them.

On the other side, proponents of open data perceive informed consent (and controlled data access in general) as a barrier to timely research. They maintain that open data leads to more research output,

and increases the probability of a scientific breakthrough. While this is true, individuals that accept to share their data openly enjoy certain privileges that others may not be able to afford themselves. They generally possess good social or economic status that protects them from the potential risks of exposing current or future health vulnerabilities [2]. In order to combat this, and to widen the subject scope to include people from varied backgrounds, we need to offer valid and substantial protection to vulnerable populations (minorities, children, sick, etc.) against abuse. Such protection could be offered through novel laws and policies (a new example is GDPR, but its impact is yet to be tested), or through health insurance and compensation to affected participants.

Furthermore, the ownership of these intellectual properties produced from the collected data, and the idea of benefit sharing them with participants, are issues of high relevance that are rarely discussed in literature [90]. While it is common practice to be compensated for research participation (generally only limited compensation), one issue that arises is the lack of benefit after medical research is successfully completed. Big data collected from participants has the potential to develop valuable products, yet subjects do not enjoy any return on investment. As of now, participants do not benefit from the intellectual properties introduced (e.g. drugs developed or medical procedures validated). Proponents of open data (even those that call for obligatory data sharing for the benefit of society) do not guarantee that the generated products (such as medicine) will actually be accessible and within reach to the public (or even to the participants, should they need it). Thus, in order to encourage subject participation (and appeal to the subjects' altruistic side), the discussion on the participation's benefit to society should widen to encompass not only knowledge advancement, but also benefit sharing and access to treatments.

The objective of this article is to further the dialogue about informed consent in the context of biomedical research. In sum, it highlights the struggles in maintaining informed consent in the age of big data. Further research is still needed to determine the best way forward with the notion of informed consent. Future research must engage individuals into designing consent mechanism that best suit them. It must also explore whether/when an exemption from consent can be practiced (legally and ethically), and the technological measures required for the protection of participants in such instances.

Acknowledgment

This work is supported by the UAE University research grant UPAR #31T121, UAE. We would like to thank the anonymous reviewers for their helpful comments/suggestions on earlier drafts of the manuscript.

References

- [1] Institutional Review Board. Required components of informed consent. <https://www.irb.cornell.edu/forms/consent.htm>, Accessed date: 6 December 2018.
- [2] Choudhury S, Fishman JR, McGowan ML, Juengst ET. Big data, open science and the brain: lessons learned from genomics. *Front Hum Neurosci* 2014;8.
- [3] Rhodes R. In defense of the duty to participate in biomedical research. *Am J Bioeth* 2008;8:37–8.
- [4] Rennie S. Regarding research participation as a moral obligation: who shoulders the burdens and who reaps the benefits? *Asian Bioeth Rev* 2010;2:308–21.
- [5] Ioannidis JP. Informed consent, big data, and the oxymoron of research that is not research. *Am J Bioeth* 2013;13:40–2.
- [6] Green DST, MacKenzie CR. Nuances of informed consent: the paradigm of regional Anesthesia. *HSS J* 2007;3:115–8. <https://doi.org/10.1007/s11420-006-9035-y>.
- [7] 1665, Duke's laws. <http://www.llmc.com/titledescfull.aspx?type=6&coll=36&div=142&set=20519>, Accessed date: 9 December 2018.
- [8] Capron AM. Where did informed consent for research come from? *The journal of law. Med Ethics* 2018;46:12–29.
- [9] Pratt V, Davis. <https://www.ravellaw.com/opinions/967d63e3e4aceb0c7ccc62e78af2d543>, Accessed date: 9 December 2018.
- [10] Schloendorff V, Soc'y of N.Y. Hospital. <https://www.ravellaw.com/opinions/13aa93a26f08eedec5264b5eaa468369?query=Schloendorff%20v.%20Soc%27y%20of%20N.Y.%20Hospital%2C%201914>, Accessed date: 9 December 2018.
- [11] Wolf SM, Clayton EW, Lawrenz F. The past, present, and future of informed consent in research and translational medicine. *J Law Med Ethics* 2018;46:7–11. <https://doi.org/10.1177/1073110518766003>.

- [12] Plant R, Pons AP. MIS Research involving human subjects: processes and IRB requirements., vol. 17 Communications of the Association for Information Systems; 2006. <https://doi.org/10.17705/1CAIS.01720>.
- [13] Group BMJP. The Nuremberg code (1947). *BMJ* 1996;313:1448. <https://doi.org/10.1136/bmj.313.7070.1448>.
- [14] WMA - The World Medical Association-WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects. <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>, Accessed date: 10 March 2019.
- [15] Haney C, Banks WC, Zimbardo PG. A study of prisoners and guards in a simulated prison. *Naval Res Rev* 1973;9.
- [16] Jones J. Bad blood: the Tuskegee syphilis experiment. *Jean Heller Broke this story in the New York times on 1972*; 1981, p. 1–8.
- [17] General Data Protection Regulation (GDPR). General Data Protection Regulation (GDPR). <https://gdpr-info.eu/>, Accessed date: 9 December 2018.
- [18] Salgo V, LELAND STANFORD JR UNIVERSITY BOARD OF TRUSTEES | FindLaw. <https://caselaw.findlaw.com/ca-court-of-appeal/1759823.html>, Accessed date: 9 December 2018.
- [19] WMA - The World Medical Association-WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects. <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>, (Accessed date: 9 December 2018).
- [20] Report Belmont. HHSGov. <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html>, Accessed date: 9 December 2018.
- [21] Federal Policy for the Protection of Human Subjects ('Common Rule'). <http://www.hhs.gov/ohrp/humansubjects/commonrule/>, Accessed date: 9 December 2018.
- [22] Simons KW, Balls M. Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the Laws, regulations and administrative provisions of the member states relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use. *Off J Eur Communities* 2001;121:34–44.
- [23] Mostert M, Bredenoord AL, Biesart MC, van Delden JJ. Big data in medical research and EU data protection law: challenges to the consent or anonymise approach. *Eur J Hum Genet* 2016;24:956.
- [24] Mittelstadt BD, Floridi L. The ethics of big data: current and foreseeable issues in biomedical contexts. *Sci Eng Ethics* 2016;22:303–41.
- [25] Church GM. The personal genome project. *Mol Syst Biol* 2005;1.
- [26] Siva N. 1000 Genomes project. Nature Publishing Group; 2008.
- [27] England G. About Genomics England. Department of Health; 2013 <https://www.genomicsengland.uk>, Accessed date: 26 February 2019.
- [28] When an Entire Country Is a Cohort | Science. http://science.sciencemag.org/content/287/5462/2398?ijkey=44d2b998408a2af32a90fb6d94ecd411d5bba6&keytype=tf_ipsecsha, Accessed date: 26 February 2019.
- [29] Precision Medicine Initiative. <http://www.nih.gov/precisionmedicine/>, Accessed date: 26 February 2019.
- [30] El Emam K. Methods for the de-identification of electronic health records for genomic research. *Genome Med* 2011;3:25.
- [31] Gelfand A. Privacy and biomedical research: building a trust infrastructure. *Biomed Comput Rev* 2012;23–8.
- [32] Cassa CA, Schmidt B, Kohane IS, Mandl KD. My sister's keeper?: genomic research and the identifiability of siblings. *BMC Med Genomics* 2008;1. <https://doi.org/10.1186/1755-8794-1-32>.
- [33] National Consumer Health Privacy Survey 2005 - California Health Care Foundation. <https://www.chcf.org/publication/national-consumer-health-privacy-survey-2005/>, Accessed date: 12 March 2019.
- [34] Greenhouse S, Barbaro M. Wal-Mart Memo suggests ways to cut employee benefit costs. *The New York Times*; 2005.
- [35] National Human Genome Research Institute (NHGRI). Genetic information and the workplace report; 2012 <https://www.genome.gov/10001732/genetic-information-and-the-workplace-report/>.
- [36] Data sharing statements for clinical trials: a requirement of the international committee of medical journal editors. *Medical Journals and Publishing | JAMA | JAMA Network*; 2017 <https://jamanetwork.com/journals/jama/article-abstract/2630773>.
- [37] Ervine C. Directive 2004/39/EC of the European Parliament and of the Council of 21 April 2004. Core statutes on company law. UK London: Macmillan Education; 2015; 757–9. https://doi.org/10.1007/978-1-137-54507-7_21.
- [38] Narayanan A, Shmatikov V. Robust de-anonymization of large sparse datasets. 2008 IEEE symposium on security and privacy. IEEE; 2008. p. 111–25.
- [39] Tocker A. Riding with the stars: passenger privacy in the nyc taxicab dataset. *Neustar Res* September 2014;15.
- [40] Samarati P, Sweeney L. Protecting privacy when disclosing information: K-anonymity and its enforcement through generalization and suppression. Technical report SRI International; 1998.
- [41] Culcane C, Rubinstein B, Teague V. Health data in an open world; 2017.
- [42] Ohm P. Broken promises of privacy: responding to the surprising failure of anonymization. *UCLA L Rev* 2009;57:1701.
- [43] Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. *Science* 2013;339:321–4. <https://doi.org/10.1126/science.1229566>.
- [44] Sweeney L, Abu A, Winn J. Identifying participants in the personal genome project by name; 2013.
- [45] Malin B. Re-identification of familial database records. AMIA annual symposium proceedings, vol. 2006 American Medical Informatics Association; 2006; 524.
- [46] Check Hayden E. Privacy protections: the genome hacker. *Nature* 2013;497:172–4. <https://doi.org/10.1038/497172a>.
- [47] Homer N, Szelinger S, Redman M, Duggan D, Tembe W, Muehling J, et al. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet* 2008;4:e1000167. <https://doi.org/10.1371/journal.pgen.1000167>.
- [48] Rodriguez-Flores JL, Fakhro K, Hackett NR, Salit J, Fuller J, Agosto-Perez F, et al. Exome sequencing identifies potential risk variants for Mendelian disorders at high prevalence in Qatar. *Hum Mutat* 2014;35:105–16. <https://doi.org/10.1002/humu.22460>.
- [49] Politou E, Alepis E, Patsakis C. Forgetting personal data and revoking consent under the GDPR: challenges and proposed solutions. *J Cybersecurity* 2018;4(1).
- [50] Appelbaum PS, Waldman CR, Fyer A, Klitzman R, Parens E, Martinez J, et al. Informed consent for return of incidental findings in genomic research. *Genet Med* 2013;16:367–73.
- [51] Paasche-Orlow MK, Taylor HA, Brancati FL. Readability standards for informed-consent forms as compared with actual readability. *N Engl J Med* 2003;348:721–6.
- [52] Williamson JML, Martin AG. Assessing the readability statistics of national consent forms in the UK. *Int J Clin Pract* 2010;64:322–9.
- [53] Beardsley E, Jefford M, Mileskhin L. Longer consent forms for clinical trials compromise patient understanding: so why are they lengthening? *J Clin Oncol* 2007;25:e13–4.
- [54] Sharp SM. Consent documents for oncology trials: does anybody read these things? *Am J Clin Oncol* 2004;27:570–5.
- [55] Borry P, Shabani M. Challenges of web-based personal genomic data sharing. *Life Sci Soc Pol* 2012;11:3.
- [56] Nishimura A, Carey J, Erwin PJ, Tilburt JC, Murad MH, McCormick JB. Improving understanding in the research informed consent process: a systematic review of 54 interventions tested in randomized control trials. *BMC Med Ethics* 2013;14:28.
- [57] McGuire AL, Beskow LM. Informed consent in genomics and genetic research. *Annu Rev Genomics Hum Genet* 2010;11:361–81. <https://doi.org/10.1146/annurev-genom-082509-141711>.
- [58] Terry SF. Don't just invite us to the table: authentic community engagement. *Genet Test Mol Biomarkers* 2013;17:443–5. <https://doi.org/10.1089/gtmb.2013.1545>.
- [59] Dresser R. Public preferences and the challenge to genetic research policy. *J Law Biosci* 2014;1:52–67. <https://doi.org/10.1093/jlb/1st001>.
- [60] Larson EB. Building trust in the power of “big data” research to serve the public good. *JAMA* 2013;309:2443–4.
- [61] Budin-Ljasec I, Teare HJ, Kaye J, Beck S, Bentzen HB, Caenazzo L, et al. Dynamic consent: a potential solution to some of the challenges of modern biomedical research. *BMC Med Ethics* 2017;18.
- [62] Dankar FK, Ptitsyn A, Dankar SK. The development of large-scale de-identified biomedical databases in the age of genomics—principles and challenges. *Hum Genomics* 2018;12:19. <https://doi.org/10.1186/s40246-018-0147-5>.
- [63] Grady Christine, Eckstein L, Berkman B, Brock D, Cook-Deegan R, Fullerton SM. Broad consent for research with biological samples: workshop conclusions. *Am J Bioeth* 2015;15:34–42.
- [64] Ruyter KW, LÖuk K, Jorqui M, Kvalheim V, Cekanaukaite A, Townend D. From research exemption to research norm: recognising an alternative to consent for large scale biobank research. *Med Law Int* 2010;10:287–313.
- [65] Caulfield T, Kaye J. Broad consent in biobanking: reflections on seemingly insurmountable dilemmas. *Med Law Int* 2009;10:85–100.
- [66] Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balsler JR, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther* 2008;84:362–9. <https://doi.org/10.1038/clpt.2008.89>.
- [67] Dankar FK, Badji R. A risk-based framework for biomedical data sharing. *J Biomed Inform* 2017;66:231–40.
- [68] Grady C. Institutional review boards. *Chest* 2015;148:1148–55.
- [69] Fletcher B, Gheorghe A, Moore D, Wilson S, Damery S. Improving the recruitment activity of clinicians in randomised controlled trials: a systematic review. *BMJ Open* 2012;2:e000496.
- [70] Wilbanks J, Friend SH. First, design for data sharing. *Nat Biotechnol* 2016;34:377.
- [71] Boutin NT, Mathieu K, Hoffnagle AG, Allen NL, Castro VM, Morash M, et al. Implementation of electronic consent at a biobank: an opportunity for precision medicine research. *J Personalized Med* 2016;6:17.
- [72] Coathup V, Teare HJ, Minari J, Yoshizawa G, Kaye J, Takahashi MP, et al. Using digital technologies to engage with medical research: views of myotonic dystrophy patients in Japan. *BMC Med Ethics* 2016;17:51.
- [73] Javaid MK, Forestier-Zhang L, Watts L, Turner A, Ponte C, Teare H, et al. The RUDY study platform—a novel approach to patient driven research in rare musculoskeletal diseases. *Orphanet J Rare Dis* 2016;11:150.
- [74] Weitzman ER, Kaci L, Mandl KD. Acceptability of a personally controlled health record in a community-based setting: implications for policy and design. *J Med Internet Res* 2009;11.
- [75] Irizarry T, Dabbs AD, Curran CR. Patient portals and patient engagement: a state of the science review. *J Med Internet Res* 2015;17.
- [76] Goetz T. 23andMe will decode your DNA for \$1,000: welcome to the age of genomics. *Wired Mag* 2007.
- [77] Vayena E, Blasimme A. Health Research with big data: time for systemic oversight. *J Law Med Ethics* 2018;46:119–29.
- [78] Whitley EA, Kanellopoulou N, Kaye J. Consent and research governance in biobanks: evidence from focus groups with medical researchers. *Public Health Genomics* 2012;15:232–42.
- [79] Steinsbekk KS, Myskja BK, Solberg B. Broad consent versus dynamic consent in biobank research: is passive participation an ethical problem? *Eur J Hum Genet* 2013;21:897–902.
- [80] Kutner M, Greenburg E, Jin Y, Paulsen C. The health literacy of America's adults: results from the 2003 national assessment of adult literacy. NCES 2006–483. National Center for Education Statistics; 2006.

- [81] Koenig BA. Have we asked too much of consent? *Hastings Center Rep* 2014;44:33–4.
- [82] Angrist M. Eyes wide open: the personal genome project, citizen science and veracity in informed consent. *Pers Med* 2009;6:691–9.
- [83] Keyserling H. *South American meditations on hell and heaven in the soul of man*. Harper & brothers; 1932.
- [84] Can Lebanon's new plan end stigma of HIV/AIDS? | News | Al Jazeera. <https://www.aljazeera.com/news/2015/12/lebanon-plan-stigma-hiv-aids-151202083548675.html>; 2015. Accessed date: 12 March 2019.
- [85] Grime002. Seeing beyond the margins: challenges to informed inclusion of vulnerable populations in research. consortium on law and values. <https://consortium.umn.edu/publications/seeing-beyond-margins-challenges-informed-inclusion-vulnerable-populations-research>; 2018. Accessed date: 7 March 2019.
- [86] McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, Larson EB, et al. The eMERGE network: A consortium of biorepositories linked to electronic medical records data for conducting genomic studies; 2011.
- [87] Leitsalu L, Metspalu A. From biobanking to precision medicine: The Estonian experience. *Genomic and precision medicine*. 3rd ed. Elsevier; 2016. p. 119–29.
- [88] Munung NS, Marshall P, Campbell M, Littler K, Masiye F, Ouwe-Missi-Oukem-Boyer O, et al. Obtaining informed consent for genomics research in Africa: analysis of H3Africa consent documents. *J Med Ethics* 2016;42:132–7. <https://doi.org/10.1136/medethics-2015-102796>.
- [89] The Faroe Genome Project selects 10x Genomics' Linked-Reads to sequence the Faroese Population | Business Wire, <https://www.businesswire.com/news/home/20171010005689/en/Faroe-Genome-Project-selects-10x-Genomics%E2%80%99-Linked-Reads> (accessed November 26, 2018).
- [90] Tene O, Polonetsky J. Big data for all: Privacy and user control in the age of analytics. *Nw J Tech & Intell Prop* 2012;11(5):239 [11:xxvii].
- [91] Susan M. Wolf, Emily Scholtes, Barbara A. Koenig, Gloria M. Petersen, Susan A. Berry, Laura M. Beskow, Mary B. Daly, Conrad V. Fernandez, Robert C. Green, Bonnie S. LeRoy, Noralane M. Lindor, P. Pearl O'Rourke, Carmen Radecki Breitkopf, Mark A. Rothstein, Brian Van Ness, Benjamin S. Wilfond, 2018, Pragmatic Tools for Sharing Genomic Research Results with the Relatives of Living and Deceased Research Participants, <https://doi.org/10.1177/1073110518766024> (accessed March 7, 2019).
- [92] Prainsack B, Buyx A. A solidarity-based approach to the governance of research biobanks. *Med Law Rev* 2013;21:71–91.
- [93] Pullman D, Etchegary H, Gallagher K, Hodgkinson K, Keough M, Morgan D, et al. Personal privacy, public benefits, and biobanks: a conjoint analysis of policy priorities and public perceptions. *Genet Med* 2012;14:229.
- [94] Crawford K, Miltner K, Gray ML. Critiquing big data: politics, ethics, epistemology. *Int J Commun* 2014;8 19328036.
- [95] Horowitz CR, Robinson M, Seifer S. Community-based participatory research from the margin to the mainstream: are researchers prepared? *Circulation* 2009;119: 2633–42.
- [96] Sweeney L. K-anonymity: a model for protecting privacy. *International journal of uncertainty. Fuzziness Knowl Based Syst* 2002;10:557–70.
- [97] Machanavajjhala A, Kifer D, Gehrke J, Venkatasubramanian M. l-diversity: privacy beyond k-anonymity. *ACM Trans Knowl Disc Data (TKDD)* 2007;1:3.
- [98] Dwork C. Differential privacy: a survey of results. Proceedings of the 5th international conference on Theory and applications of models of computation. Berlin, Heidelberg: Springer-Verlag; 2008. p. 1–19.
- [99] Domingo-ferrer J, Sebe F, Castella-Roca J. On the security of noise addition for privacy in statistical databases. *Privacy in statistical databases 2004*. Springer; 2004. p. 149–61.
- [100] Surendra H, Mohan HS. A review of synthetic data generation methods for privacy preserving data publishing. *Int J Sci Technol Res* 2017;6.
- [101] El Emam K, Dankar FK, Neisa A, Jonker E. Evaluating the risk of patient re-identification from adverse drug event reports. *BMC Med Inform Decis Mak* 2013;13:114.
- [102] Armando A, Bezzi M, Metoui N, Sabetta A. Risk-aware information disclosure. In: Garcia-Alfaro J, Herrera-Joancomarti J, Lupu E, Posegga J, Aldini A, Martinelli F, et al, editors. *Data privacy management, autonomous spontaneous security, and security assurance*. Springer International Publishing; 2015. p. 266–76.
- [103] Badji R, Dankar FK. A risk-aware access control model for biomedical research platforms. *ICISSP 2018*:322–8.
- [104] Ohmann C, Canham S, Banzi R, Kuchinke W, Battaglia S. Classification of processes involved in sharing individual participant data from clinical trials. *F1000Research* 2018;7.
- [105] Ohmann C, Banzi R, Canham S, Battaglia S, Matei M, Ariyo C, et al. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. *BMJ Open* 2017;7:e018647.
- [106] El Emam K. Risk-based De-identification of health data. *IEEE Secur Priv* 2010;8. <https://doi.org/10.1109/MSP.2010.103> 64–7.
- [107] Unlocking big data for healthcare-Privacy Analytics, White paper; 2016 <https://privacy-analytics.com/de-id-university/white-papers/unlocking-big-data-healthcare/>, Accessed date: 9 March 2019.
- [108] Dwork C, Mcsherry F, Nissim K, Smith A. Calibrating noise to sensitivity in private data analysis. *Lecture notes in computer science*. Springer; 2006. p. 265–84.
- [109] Kairouz P, Oh S, Viswanath P. Secure multi-party differential privacy. *Adv Neural Inf Process Syst* 2015:2008–16.
- [110] Dankar FK, Boughorbel S, Badji R. Using robust estimation theory to design efficient secure multiparty linear regression. *Proceedings of the 2016 joint EDBT/ICDT workshops*; 2016.
- [111] El Emam K, Samet S, Hu J, Peyton L, Earle C, Jayaraman GC, et al. A protocol for the secure linking of registries for HPV surveillance. *PLoS ONE* 2012;7:e39915. <https://doi.org/10.1371/journal.pone.0039915>.
- [112] Dankar F. Privacy preserving linear regression on distributed databases. *Trans Data Priv* 2015;8:3–28.
- [113] Yang Z, Zhong S, Wright RN. Privacy-preserving queries on encrypted data. *Computer security-ESORICS 2006*. Springer; 2006. p. 479–95.
- [114] Bonte C, Makri E, Ardeshirdavani A, Simm J, Moreau Y, Vercauteren F. Privacy-preserving genome-wide association study is practical *Cryptology ePrint Archive*; 2017.
- [115] Song DX, Wagner D, Perrig A. Practical techniques for searches on encrypted data. *IEEE symposium on security and privacy*; 2000. p. 44–55.
- [116] El Emam K, Arbuckle L, Essex A, Samet S, Wang L. Secure surveillance of antimicrobial resistant organism colonization or infection in Ontario. *Plos One* 2014;9(4): e39915.
- [117] Dankar FK, Madathil N, Dankar SK, Boughorbel S. Secure multi-party computations for biomedical data using distributed statistical learning. *JMIR Med Inf* 2019. <https://doi.org/10.2196/12702>.
- [118] Sutrop M. How to avoid a dichotomy between autonomy and beneficence: from liberalism to communitarianism and beyond. *J Intern Med* 2011;269: 375–9.