IIC (2019) 50:1052–1081
https://doi.org/10.1007/s40319-019-00873-2

ARTICLE

Towards a Paradigm Shift in Governing Data Access and Related Intellectual Property Rights in Big Data and Health-Related Research

Pamela Andanda

Published online: 15 October 2019

Abstract Big data is revolutionising the health care sector by making it easier to analyse large volumes of data. This enables health care providers to monitor individuals or systems in real time. However, the main concern with regard to big data in biomedicine is how to overcome the barriers to using such data for health-related research. The landscape of big data is still evolving and the law has not developed suitable principles for governing access to big data. This paper sketches the fuzzy contours of data ownership and related intellectual property rights to demonstrate that ownership is a concept that is ill-suited for governing rights in big data. The dawn of big data calls for an alternative normative framework. This framework must be capable of reconciling competing societal, individual and industries' interests in the data with a view to ensuring fair access while minimising legal and ethical risks. Ultimately, the paper proposes a paradigm shift from ownership to custodianship in the governance of access and use of big data, particularly in international health-related research.

Keywords Big data · Custodianship · Data access · Health data · Health research · Intellectual property

An earlier version of this paper was presented at my inaugural lecture at the University of the Witwatersrand, Johannesburg on 4 October 2017.

P. Andanda (✉)
Professor of Law, University of the Witwatersrand, Johannesburg, South Africa
e-mail: Pamela.Andanda@wits.ac.za

1 Introduction

One of the benefits of using big data in the health care sector is the prospect of improving the efficiency of service delivery.[1] The efficiency lies in the potential of analysing large volumes of data, which enables health care providers to monitor individuals or systems in real time.[2] However, there are concerns about sharing and reusing data in formats that include big data practices.[3] The barriers mostly stem from privacy and security regulations as well as legal concerns,[4] which can deepen the existing inequalities between peoples and countries.[5] Additionally, the main concern with regard to big data in biomedicine is how to overcome the barriers to sharing and reusing such data for health-related research.[6]

To contextualise the above concerns, it is worth considering the nature of big data briefly. Although there is no precise definition of big data, its attributes are well documented in current literature. These are comprised of the large increase in the *volume* of data that can be generated and stored, the *velocity* with which data can be delivered to foster decision-making in real time, the *variety* of formats in which data can be adopted, the *veracity* or confidence level that is associated with certain types of data[7] and the ability to extract *value* by "identifying what is valuable and then transforming and extracting [it] for analysis".[8] Accordingly, the term is defined by reference to five attributes, namely volume, velocity, variety, veracity and value.[9]

The attribute of variety, which highlights the diverse interests, is evident from the following sources of information that make up big data in health-related research: "electronic health care records, social media, patient summaries, genomic and pharmaceutical data, test results, claims, telemedicine, mobile apps, home monitoring, clinical trials, sensors and information on wellbeing, behaviour and socioeconomic indicators".[10]

The diverse sources of big data show how the line between health care and health research has become blurred.[11] The scope of data producers has equally expanded with the growing numbers of citizens and citizen scientists with access to mobile cellular networks.[12] These developments have led to proposals that advocate outright ownership of data by data subjects[13] or monetisation of their access and

[1] Mellado (2015).

[2] The IET and Royal Academy of Engineering (2015).

[3] Mattioli (2014); Borgman (2012).

[4] Bender (2015).

[5] UNESCO (2017); UNESCO (2015).

[6] *See* Borgman (2012), who observes that data sharing is not taking place as expected.

[7] Luna et al. (2014); Mogha et al. (2013); Laney (2001).

[8] Dijcks (2013), p. 4.

[9] Ward and Barker (2013); Dijcks (2013).

[10] European Commission (2014), p. 4.

[11] UNESCO (2017).

[12] ITU World Telecommunication (2016).

[13] Kish and Topol (2015).

control rights.[14] The proposals raise three fundamental questions; are data capable of being owned? Who owns data? And what is the basis of such ownership?

Different approaches to resolving the above questions have been indecisive[15] and current literature has mostly focused on technical, privacy and security issues as the main challenges of implementing big data.[16] For example, Wyber and colleagues have observed that the field of big data "is fraught with ethical, regulatory and technological issues" and have accordingly called for a "move from a reactive model to a proactive, norm-forming approach" in global health governance.[17] Zwitter has also observed that "a rethinking in philosophy, professional ethics, policy-making, and research" is essential in the era of big data.[18] The European Commission has suggested steps that can be taken in this regard by calling for flexibility of stakeholders' roles and responsibilities in order to avoid single actor responsibilities across the data value chain.[19] This paper focuses on how claims of data ownership are impacting data sharing and implementation of big data in health-related research. The point of departure is UNESCO's observation that the concept of ownership is no longer an adequate normative framework in the era of big data.[20]

The landscape of big data is still taking shape and this paper attempts to contribute to the much-needed legal and policy guidance by examining the fuzzy contours of data ownership and related intellectual property rights (IPRs), which have been cited as the main obstacles or possible solutions to data sharing. Notably, issues of data ownership remain mostly unresolved amidst calls for regulating data as property.[21] The central argument in this paper is that ownership is a concept that is ill-suited for governing rights in big data. The dawn of big data calls for an alternative normative framework. This framework must be capable of reconciling competing societal, individual and industries' interests in the data with a view to ensuring fair access while minimising legal and ethical risks, as recommended by the OECD.[22] The ultimate aim of the paper is to propose a paradigm shift from ownership to custodianship in the governance of big data, particularly in international health-related research. The focus on international health-related research is warranted by the fact that the digital nature of big data has made research more globalised and collaborative, yet few countries have developed suitable policies or strategies to govern the use of big data in the health sector.[23]

[14] Hall (2010).

[15] Hoeren (2014).

[16] Luna et al. (2014); Wyber et al. (2015).

[17] Wyber et al. (2015), pp. 205 and 206.

[18] Zwitter (2014), p. 1.

[19] European Commission (2016), p. 44.

[20] UNESCO (2017).

[21] Ritter and Mayer (2018).

[22] OECD (2017).

[23] World Health Organisation (2016).

Issues that are related to access to information and data-owning companies' concerns about disclosure of IPRs in data that they own have not been considered sufficiently.[24] In contrast, a lot of valuable research has been done on stakeholders' perspective on data sharing of public health research by LMICs.[25] This paper seeks to fill the gap in the literature by focusing on two specific regulatory issues, namely ownership of data and related IPRs that data holders rely on to impede sharing and reuse of data. The second part provides an overview of the concerns in health-related research using big data as well as ownership and IPR issues that impede data sharing. The third part discusses the proposed paradigm shift from ownership to custodianship and paves the way for an explanation, in the fourth part, of the attributes of the proposed alternative normative framework for governing competing rights in big data.

2 Concerns in Health-related Research Using Big Data

The increased use of mobile devices and wearable or implanted devices and biosensors, which produce, collate and facilitate access to data have led to a wider circulation of digital health data.[26] The production and distribution of health data through these means effectively disrupt the conventional modes of data collection through established institutional channels such as hospitals[27] and increase the number of people who are involved as research participants.[28] Health data may be accessed directly from devices or may be voluntarily shared by individuals on social media. The real-life data that are derived from these sources can be accessed and used for research in ways that raise ethical issues related to ownership and data access.[29]

Before discussing the concerns in health-related research, it is important to clarify the different types of data sets and the interests that are at stake. The first set is real-life data, which can be obtained from wearable digital devices and social media platforms where individuals share their health data from these devices with their peers. In this context, device and social media users co-produce and circulate the data.[30] Three parties are typically interested in this type of data, namely device or platform users who share data with each other, researchers who use the data to support their theories, and businesses that have vested financial interests in analysing the data.[31] Notably, researchers and businesses that analyse the data can incur considerable investments in generating a second type of data set of high quality. From the sources of health data that were mentioned in the introduction to

[24] Megget (2011).

[25] Denny et al. (2015); Hate et al. (2015); Merson et al. (2015).

[26] Erikainen et al. (2019).

[27] Kallinikos and Tempini (2014).

[28] Erikainen et al. (2019).

[29] Erikainen et al. (2019); Ostherr et al. (2017).

[30] Erikainen et al. (2019), p. 2.

[31] Ostherr et al. (2017).

this paper, this second type of data set can consist of genomic, pharmaceutical data or clinical trial results. An interesting observation is that these types of data sets may be derived from individual level data with implications on the rights of data subjects. Both data sets are the objects of contested access and claims of ownership, particularly when producers of high quality data sets attempt to protect their interests through trade secrecy and mechanisms that can guarantee exclusive rights, as further explained in Sect. 2.3 of this paper.

Health-related research thrives on data sharing from diverse sources since it is highly interdisciplinary in nature.[32] Additionally, the research is mostly conducted in a globalised and collaborative context. In the era of big data, sharing of digital data occurs on a global scale. Data sharing is vital because data generators, analysts and researchers have to work together as a team for purposes of making appropriate use of big data.[33] The unique circumstances, particularly in LMICs that have implications on the uptake of big data in health-related research, are the large size of the population and the complexity of health care delivery, which have led to a gap between health care delivery and population health.[34] Wyber and colleagues argue that this gap may be bridged and health care outcomes can be improved using the big data approach.[35] This hope can only be realised if three concerns in health-related research using big data are addressed appropriately. These are: access to data, data subjects' consent to data processing for research, and ownership claims that can impede data sharing.

2.1 Data Access

With the emergence of big data, the risk of "putting so much personal data in the hands of either companies or governments" is real and this can lead to misuse of such data.[36] One of the critical questions in this regard is access,[37] which is closely related to the data subjects' control over who may access, use and share their information. Companies may also be interested in timely access and dissemination of the data in a manner that provides investment incentives to stakeholders such as firms that collect and process data.[38] Catering for these interests requires other mechanisms, such as legislative provisions, oversight mechanisms, and procedures for the use of health data,[39] to be in place to foster control. Accordingly, Pentland and colleagues have suggested ensuring data access as one of the means of supporting the development of big data health systems. In their view, this entails updating "privacy and data ownership policies to ensure that data are accessible to

[32] van Panhuis et al. (2014).

[33] Nicen (2015).

[34] Wyber et al. (2015), p. 203.

[35] *Ibid.*

[36] Pentland et al. (2013), p. 5.

[37] Boyd and Crawford (2012), p. 664.

[38] European Commission (2019).

[39] Vayena and Blasimme (2017), p. 503.

patients and their healthcare providers".[40] The element of control should therefore be understood as a way of building trust in health-related research, thus encouraging data subjects to agree to their data being made accessible in a manner that safeguards their interests.

A qualitative study on five LMICs[41] established that while most stakeholders are open to sharing health research data, they have concerns about ownership and allowing free access to data. Additionally, data sharing is a challenge in LMICs and at a global level due to a lack of guidance and regulations.[42] It is worth noting that there are no "locally enforceable data protection rules and standards" in LMICs.[43] Nevertheless, stakeholders in these countries are comfortable to share de-identified data for academic and public health purposes so long as anonymity of the research participants' personal information is guaranteed but not beyond these limits.[44] For instance, some stakeholders argue that "making data available [for re-use] actually demonstrates respect for the respondents, in that you care about what they're saying, it's not just something that you use and discard".[45] This shows that making data accessible in a manner that respects the wishes of data subjects can build trust and encourage them to make their data accessible for health-related research. In this regard, Vayena and Blasimme have correctly argued that "the availability of data control – being a sign of respect for people's interests – may promote rather than hinder the propensity to share data for health and health research related purposes".[46]

Data exportation and re-use for commercial purposes, on the other hand, are perceived as a threat to the local researchers and communities since there is no guarantee of local benefits[47] and consequently a threat to the local researchers and participants.[48] The following statement from one of the stakeholders is very instructive on the issue of data sharing in a manner that is beneficial to the local researchers and communities:

> there has to be a benefit sharing component that's in the data sharing process and the benefit sharing has to be … done in a critical way where there is *not just benefit for the investigator who is now going to have a patent and generating billions versus the community who's still living in poverty*.[49]

[40] Pentland et al. (2013), p. 2.

[41] India, Thailand, Vietnam, South Africa, and Kenya; *see* Denny et al. (2015).

[42] Denny et al. (2015); Parker and Bull (2015).

[43] Bellagio Big Data Workshop Participants (2014), p. 32.

[44] Denny et al. (2015), p. 294; Hate et al. (2015), p. 242.

[45] LMIC research manager, quoted in Denny et al. (2015), p. 298.

[46] Vayena and Blasimme (2017), p. 504.

[47] Merson et al. (2015), p. 256; *see also* study by Shah et al. (2018), which established that for unknown reasons European research participants prefer to share data with universities and are least happy to share with commercial companies.

[48] Denny et al. (2015), p. 297.

[49] LMIC research manager, quoted in Denny et al. (2015), p. 298. Emphasis added.

An additional challenge that makes data exportation and re-use for commercial purposes to be perceived as a threat is a lack of legal capacity in LMIC-based research institutions to ensure that the agreements that they enter into are equitable enough to cover issues such as fair data ownership, IPRs and future benefit sharing.[50] This observation resonates with Indian stakeholders' wish that recompense be expressed "more in terms of benefits to communities than in the form of acknowledgment or authorship".[51]

The urge to ensure benefits for data subjects can cause fear of loss of control, based on the inability to control the nature of use and beneficiation to the local communities by secondary end users. This concern has been expressed particularly when data are shared with developed country partners in circumstances where research data are handed over for rapid analysis in developed countries with technological and technical capacity.[52] In research involving big data, the concern also relates to the possibility of incidental findings, which may have limited clinical relevance due to the scale of the research.[53] The concern is essentially linked to the question of custodianship over data, which can be challenging when data are used on a global scale, thus making it difficult for data subjects to have control over their data.[54]

Informed consent, as discussed in the next subsection of the paper, allows data subjects to maintain control over the use of their data. The conditions under which informed consent is given by data subjects usually enable them to determine whether or not their expectations and best interests are taken into account.[55] Furthermore, the digital world in which data are used presents threats of data subjects losing control over their data.[56] The era of big data thus makes it difficult for data subjects to foresee specific future uses and users, mostly due to the complex interrelationships between multiple and changing data sources.[57] Therefore, while consent may take care of concerns related to the nature of use to some extent, it does not sufficiently address the issue of beneficiation to the local communities by secondary end users. It is in this context that LMIC stakeholders have suggested that "additional regulations to protect the community's interests should be applied to non-local data-access requests".[58] In Vietnam, for instance, stakeholders have suggested that international data sharing policies, mostly developed by funders and publishers, "should not be imposed without consideration of local research culture, needs, and expectations".[59]

[50] Sankor and Ijsselmuiden (2011); Sack et al. (2009).

[51] Hate et al. (2015), p. 246.

[52] Denny et al. (2015), p. 297.

[53] Lipworth et al. (2017).

[54] *Ibid.*

[55] Vayena and Blasimme (2017).

[56] *Ibid.*

[57] UNESCO IBC (2017), para. 51.

[58] Denny et al. (2015), p. 296.

[59] Merson et al. (2015), p. 252.

The above suggestions for averting a potential loss of control over data are difficult to implement outside the health care and research settings. For example, commercial health-related databases and data collected from social networking platforms or commercial apps that encourage data subjects to upload their data may be difficult to control.[60] Data subjects that upload data in this manner may include citizen scientists whose consent should determine the terms of using data.[61] However, the possibility of foreseeing and specifying terms of use through consent has become a challenge with advances in big data and citizen science,[62] which entails citizens becoming "experimenters, stakeholders and purveyors of data".[63] These challenges have led to calls for a move beyond consent to a broader framework of accountability, which reckons with issues such as harm and risk assessment.[64]

Recent developments have introduced a number of rights to ensure that data subjects maintain some control over their personal data. For example, the European Union's (EU) General Data Protection Regulation (GDPR) provides for rights to access, rectification, erasure and data portability.[65] Notably, Recital 63 of the GDPR also protects third-party rights by specifying that data subjects' rights to access should not adversely affect the rights of others to trade secrets or IP such as copyright that protects software. Recital 156 provides for derogations, *inter alia*, in the public interest, and for scientific or historical research purposes, provided that conditions and safeguards are in place to protect data subjects' rights. In addition, data processing should be pursuant to proportionality and necessity principles. Data processing for scientific purposes should also comply with other legislation.

The right to portability entitles data subjects to *receive* their personal data "in a structured, commonly used and machine-readable format and *have the right to transmit* those data to another controller".[66] This has two important implications on access and sharing of health data. Firstly, data subjects are entitled to receive their complete records and, secondly, they can freely share or transfer the data to any person that they wish, thus fostering data interoperability, competition and accessibility.[67] If these effects are achieved in practice, then this requirement can address concerns about loss of control over data, thus facilitating sharing of data

[60] UNESCO IBC (2017); Vayena and Blasimme (2017).

[61] Cheung (2018).

[62] *Ibid.*

[63] *See* Opinion No. 29 of the European Group on Ethics in Science and New Technologies to the European Commission: The Ethical Implications of New Health Technologies and Citizen Participation, available at http://ec.europa.eu/research/ege/pdf/opinion-29_ege.pdf, para. 1.2.2, where these concepts are defined, respectively, as "patients participating in various degrees in experimentation", "patient expert groups" and "citizens/patients sending data through ICT, mobiles, digital devices".

[64] Cheung (2018).

[65] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *See* Arts. 15, 16, 17 and 20, respectively.

[66] GDPR, Art. 20(1) (emphasis added).

[67] Vayena and Blasimme (2017), p. 508; Blasimme et al. (2018).

with data subjects' health care providers.[68] Indeed there are settings that warrant imposing the duty to ensure data access and data interoperability.[69] Notably, health-related research warrants this approach due to the sensitivity of the data and the need to serve the common good as envisaged by the data subjects' intention in consenting to the sharing of their data for research purposes.

The GDPR is an EU regulation without an automatic extraterritorial application. However, Art. 3 provides for three instances of extraterritorial application. Firstly, the regulation applies to controllers or processors that are established in the Union, irrespective of whether the processing takes place in the Union or not. Secondly, it applies to controllers or processors that are not established in the Union if the processing involves monitoring the behaviour of data subjects in the Union. Thirdly, it applies to controllers or processors that are not established in the Union if Member State law applies by virtue of public international law. Article 50 also provides for international cooperation between the EU and third countries in the enforcement of privacy laws. In this case, the GDPR would apply if the research is funded by the EU or if appropriate mechanisms are in place for international cooperation with third countries.

2.2 Data Subjects' Informed Consent

Obtaining data subjects' consent ensures the preservation of their autonomy to intervene in the decision-making process regarding the use of their data.[70] Consequently, attempts have been made to use informed consent in governing the use of data by specifying what should be done with individual research data and the terms of the contractual obligations that are considered "to guard privileged information on behalf of the research funder or sponsors".[71]

Using data from sources such as social media raises additional issues that underpin big data as a phenomenon. Boyd and Crawford, have accurately observed that social media users create their data in highly context-sensitive spaces and they are unlikely to permit their data to be used elsewhere; many of these users "are not aware of the multiplicity of agents and algorithms currently gathering and storing their data for future use"; researchers are rarely part of the users' imagined audience; and "users are not necessarily aware of all the multiple uses, profits, and other gains that come from information they have posted".[72] The two authors have raised the following pertinent questions[73]:

> Should someone be included as a part of a large aggregate of data? What if someone's "public" blog post is taken out of context and analyzed in a way that the author never imagined? What does it mean for someone to be

[68] Rumbold and Pierscionek (2017).

[69] European Commission (2019), p. 9.

[70] Council of Europe (2017), para. 7.1.

[71] Denny et al. (2015), p. 297.

[72] Boyd and Crawford (2012), p. 673.

[73] *Ibid*., p. 672.

spotlighted or to be analyzed without knowing it? Who is responsible for making certain that individuals and communities are not hurt by the research process? What does informed consent look like?

The above questions demonstrate that using data for health-related research requires the issue of informed consent to be addressed appropriately as provided in Art. 6 of the Universal Declaration on Bioethics and Human Rights (UDBHR), which requires scientific research to be carried out "with the prior, free, express and informed consent of the person concerned".[74] At the point of obtaining informed consent from data subjects, they have the right to know if access to the data that they are contributing will be open or limited due to commercial reasons.[75] Providing such relevant information to data subjects allays their fears regarding the possible misuse of their data in the course of commercialisation, particularly in view of the fact that "data sharing is not yet commonplace and trust in such processes is established slowly".[76] In the context of big data, there are calls for a suitable and dynamic model of informed consent that can facilitate access and respect of the data subjects' autonomy,[77] mostly because health information privacy laws are rather permissive and patient-generated information is not governed by privacy laws.[78] Additionally, using technical measures such as anonymisation cannot always avert possible re-identification.[79]

Considering the diverse sources of big data and the fact that the future utility of big data is usually uncertain at the point of obtaining informed consent, Mittelstadt and Floridi have correctly argued that consent cannot be truly informed because of the difficulties of predicting and informing the data subjects of the future uses and consequences of the data.[80] Evidently, the rapid production, collection, use and sharing of health-relevant data in the era of big data are challenging the extent to which informed consent can be used to preserve data subjects' autonomy, thus calling for new models of consent in health-related research.

Three models of consent that have emerged in the context of big data are broad consent, opt-out consent and dynamic consent.[81] In broad consent, data subjects consent to "a range of possible research that could be done with [their] information in relation to a specific area or line of investigation".[82] For broad consent to be valid, relevant forms of governance and safeguards, such as relevant review committees that ensure the protection of data subjects' rights, must be in place.[83] The opt-out model assumes that data can be used unless the data subject has

[74] UNESCO (2005).

[75] Alter and Vardigan (2015), p. 318.

[76] *Ibid*., p. 319.

[77] UNESCO (2017).

[78] Thorpe and Gray (2015).

[79] Tene and Polonetsky (2013), p. 251.

[80] Mittelstadt and Floridi (2016), p. 312.

[81] UNESCO IBC (2017); Vayena and Blasimme (2017).

[82] UNESCO IBC (2017), para. 52.

[83] UNESCO IBC (2017), para. 52; Steinsbekk et al. (2013).

explicitly opted out. The challenge with this model is that data subjects may not be adequately informed of the terms of use, particularly in commercial data bases or the use of social media. The dynamic model allows data subjects to update their consent on an ongoing basis. Although this model has mostly been used in biobanking it can also be used in circumstances that entail multiple and varied uses of data, such as big data, where different kinds of consent may be required over time.[84] It provides an open communication process between data subjects and researchers thus ensuring that the evolving data subject preferences are taken into consideration in the adaptive process, which gives them more control over their data for the duration of the research.[85]

Ideally, a suitable consent model should engage with the individuals beyond the point of data collection, thus allowing them to harness big data for their own personal use and to constrain unacceptable uses.[86] Such a model can only work within the context of custodianship, which recognises big data as a common good for the benefit of humankind. It thus entails granting individuals "meaningful rights to access their data in a usable, machine-readable format" while at the same time striking a delicate balance between providing insight into the decisional criteria of organisations that draw conclusions from personal information and the protection of IPRs.[87]

The issue of consent, as highlighted above, shows that competing interests can be governed better if researchers and other stakeholders focus on acting ethically and are accountable in addressing each other's concerns.[88] Therefore, issues of consent and accountability can be better governed through an alternative normative framework that takes into account the diverse stakeholders' interests in the data.

2.3 Ownership of Big Data and Its Correlation with IPRs that Impede Data Sharing, Reuse and Accountability

The concept of ownership can refer to the right to "control" data or the right to "benefit from" data.[89] The right to control access to data directly affects data sharing and is closely linked to the conventional use of the term in IP law where claims to ownership of intellectual content must be recognised by law for the rights to be effective.[90] Notably, the terms "data controller" or processor are used in data protection law to denote, respectively, the person, entity or body that "determines the purposes and means of the processing of personal data" and "processes personal data on behalf of the controller".[91] This, however, leaves the issue of ownership ill-

[84] Kaye et al. (2015).

[85] Vayena and Blasimme (2017).

[86] Tene and Polonetsky (2013), p. 242; Innes (2010).

[87] Tene and Polonetsky (2013), pp. 242–243 and 271.

[88] Boyd and Crawford (2012), p. 672.

[89] Mittelstadt and Floridi (2016), p. 319.

[90] Bettig (2018).

[91] *See* for example the GDPR, Art. 4.

defined.[92] The right to benefit from data relates more to custodial rights such that data custodians are expected to allow data subjects to access and utilise the data for their own benefit.[93] The first formulation (control of access) is accordingly used in this section of the paper and it should be understood as a controversial concept in the context of big data. This is because the underlying information or data, over which IP law seeks to create related property rights, are the building blocks of science and the results of multiple producers' efforts.[94]

So far, the issue of who owns data generally or even in the context of big data has not been explored sufficiently in current literature.[95] The complexity of the issue lies in the public-good character of such data and the diverse sources of information that constitute the data. This is not surprising since the main legal barrier to sharing data, which has been identified in current literature, is based on ownership of data insofar as those who "collect public health data are also often responsible for the protection of individual and community privacy and may feel that a guardianship or ownership role is bestowed on them by the public".[96] To further illustrate the complexity of the issue, Burtscher and Fritz have likened big data to a block of marble that a number of cutters are working on such that "various legal concepts including data privacy, database rights, IP rights, antitrust law as well as the basic civil rights of ownership and possession are playing a role when dealing with the legal alien big data but are each only addressing bits of it".[97] The two authors have accordingly wondered whether the concept of ownership correctly captures big data in legal terms. Hoeren has also observed that the question of how "new property right" in data fits into the existing property law framework remains to be solved.[98]

The discussions in this section will demonstrate that granting property rights in data as such would be ill-advised and lacks a sound legal basis. Evidently, data are subject to access rights and restrictions but these are not property rights. For example, contracts or competition law may be used to regulate access to data but they do not create property rights.[99] These mechanisms of governing access to data are important because competitors are interested in using data to develop innovative services and products.[100]

There have also been attempts to govern issues of ownership through informed consent in the mistaken belief that data subjects own their data as aptly stated below:

[92] Blasimme et al. (2018).

[93] Mittelstadt and Floridi (2016), p. 319; *see also* Tene and Polonetsky (2013), p. 242.

[94] Haunss and Shadlen (2009), pp. 1 and 3.

[95] Denny et al. (2015); Burtscher and Fritz (2015); Hoeren (2014). Determann (2018) has extensively discussed the issue and concluded that no one owns data. However, there are few publications on this topic in the context of big data.

[96] van Panhuis et al. (2014), online.

[97] Burtscher and Fritz (2015), online.

[98] Hoeren (2014), p. 754.

[99] *See* Determann (2018), who also embraces this view.

[100] European Commission (2019), p. 73.

> whether the informed consent form [authorises] the transfer of those rights from the participant to the investigator or the sponsor – if they [the participant] have not … agreed to transfer their rights of the data [then] neither the sponsor, nor the investigator, [nor] the collaborator outside of the institution can actually say that they own the data.[101]

According to the above statement, the researchers "should not be viewed as owning the data but rather as having custodial responsibilities and rights over it".[102] They are considered to have possession of the data but not the right of ownership. This point is developed further below, in discussing copyright protection of compilations and *sui generis* rights over non-original databases. A number of stakeholders in India have, for instance, suggested that ownership and authorship rights should be clearly stipulated in the data sharing agreements while others were of the view that the organisation that had collected the data, rather than the data subjects themselves, should own the data.[103]

Burtscher and Fritz have correctly observed that "the legal discussion and legislation around allocation of ownership of anonymous data or at least of the right to use and to exclude others from using such data" is inconclusive.[104] Ekbia and colleagues have equally noted that the legal and ethical problems that changes in technology have brought to the fore have raised "new questions about the scope of individual privacy and the proper role of intellectual property protection".[105]

From the above highlights; it would appear that the prevailing view regarding ownership of data is that "the entity or individual controlling the production of anonymous data should be entitled to use them".[106] Similarly, most LMIC stakeholders are of the view that funders "reserved the right to share data because [they] possessed the intellectual property rights for that data".[107] This raises a fundamental question of what IP law protects in big data, particularly considering that the collection of data may not qualify for patent protection or even copyright protection if the information or data sets are presented in a factual manner. The compilation of big data would have to be original in the copyright law sense to qualify for protection. Ekbia and colleagues argue that "existing intellectual property laws may also need to be adapted in order to accommodate Big Data practices".[108] In the paragraphs that follow, the contents of big data that may qualify for protection under IP law are discussed with a view to determining what is protected and if such protection impedes data sharing in health-related research and should accordingly be adapted to accommodate big data practices as suggested by Ekbia and colleagues. Reference will mostly be made to the international IP

[101] Quoted in Denny et al. (2015), p. 297.

[102] Denny et al. (2015), p. 297.

[103] Hate et al. (2015), p. 245.

[104] Burtscher and Fritz (2015), online.

[105] Ekbia et al. (2015), p. 1535.

[106] Burtscher and Fritz (2015), online.

[107] Denny et al. (2015), p. 297.

[108] Ekbia et al. (2015), p. 1537.

instruments, which set the minimum standards, and regional or national laws will only be mentioned for illustrative purposes since this paper focuses on international health-related research.

Five of the sources of information that make up the contents of big data in health-related research, as mentioned in the introduction to this paper,[109] may be protected through intellectual property law if they meet the requisite requirements. These are: compilations of electronic health records, patient summaries, genomic and pharmaceutical data, test results and mobile applications. Three possible avenues of protecting these contents are copyright, *sui generis* database rights and trade secrets. The relevance of these IPRs in the context of big data and their impact on data sharing are discussed below.

2.3.1 Protection of Compilations Through Copyright

Copyright law is one means of protecting creative original compilations of data. At the international level, original structures of databases are protected under Art. 2 of the Berne Convention,[110] Art. 5 of the World Intellectual Property Organisation's (WIPO) Copyright Treaty[111] and Art. 10(2) of the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS). These international instruments protect compilations of data or other material in either machine readable format or other form. For such compilations of data or material to qualify for copyright protection, Art. 10(2) of the TRIPS Agreement provides that "the selection or arrangement of their contents [must] constitute intellectual creations". The Article further stipulates the scope of rights in such compilations by explicitly providing that "such protection, which shall not extend to the data or material itself, shall be without prejudice to any copyright subsisting in the data or material itself". Article 5 of the WIPO Copyright Treaty equally provides for the protection of compilations if the selection or arrangement of their contents constitute intellectual creations. The "protection does not extend to the data or the material itself and is without prejudice to any copyright subsisting in the data or material contained in the compilation".[112]

The fact that data as such are not protected through copyright has mistakenly been viewed as a shortcoming in copyright law.[113] Proposals have therefore been made for a new construct called "datarights" that can "be available to applicants who disclose clear and complete descriptions of their data collection and preparation methods alongside the data shaped by those methods".[114] Datarights are intended to protect data that are "collected or manipulated according to one or more methods not readily apparent to a person of ordinary skill in the art" from unauthorised use

[109] *See also* European Commission (2014), p. 4.

[110] The Berne Convention (1886).

[111] The WIPO Copyright Treaty (1996).

[112] The WIPO Copyright Treaty (1996), Art. 5.

[113] *See* Malhotra (2016), who holds such a view and advocates that data companies using trade secrets, as a means of overcoming this shortcoming in copyright law, should protect data.

[114] Mattioli (2014), p. 578.

for a limited period.[115] This subject matter essentially resembles the originality or creative skills requirement as it currently exists in the protection of original compilations through copyright except that it extends to the protection of data from unauthorised downstream use such as analysis. Notably, the proposed new construct would still leave the underlying data free for reproduction and redistribution by other stakeholders unless barred through contracts.[116] Consequently, this does not solve the alleged shortcoming in copyright law. In addition, proponents of this construct have conceded that it would not be effective in encouraging disclosure of big data practices where there are concerns related to privacy and commercial interests.[117]

Another source of confusion is the "without prejudice" clause as used in the latter parts of Art. 10(2) of the TRIPS Agreement and Art. 5 of the WIPO Copyright Treaty. The confusion arises from the fact that data or material itself, as already indicated in the first parts, is not protectable through copyright. Consequently, the clause "without prejudice to any copyright subsisting in the data or material contained in the compilation" may give the wrong impression that underlying data or material, as contained in the compilation, are protected by copyright.[118] Such misinterpretation contradicts the established copyright protection of literary and artistic works under the Berne Convention and is likely to be used to impede data sharing in the mistaken belief that data as such are capable of being subject to proprietary rights.

The protection of compilations has to be interpreted with reference to Art. 2(1) of the Berne Convention, which requires the exercise of skills in compiling the material. Accordingly, the clause should be interpreted as referring to copyright protection in literary and artistic works subsisting in the data that are included in the compilation.[119] This is, for example, clearly illustrated in the "without prejudice" wording of Art. 13 of the EU Database Protection Directive (discussed in more detail in the next section), which lists the possible rights that may subsist in the data that are included in the database.[120]

Apart from originality, copyright law requires the subject matter to be expressed in material form. Article 10(2) of the TRIPS Agreement requires the data to be compiled in machine readable or other form. Notably, the format of big data involves dynamic data sets and uses cloud computing services, which technically makes it difficult to meet this requirement.

[115] *Ibid*., pp. 578 and 581.

[116] *Ibid*., p. 579.

[117] *Ibid*., p. 580.

[118] *See* Dalal (2006), pp. 126 and 128, who incorrectly argues that the very data or material itself is protected in the form of copyright under Art. 10(2) of TRIPS.

[119] Reinbothe and von Lewinski (2015), para. 7.5.22.

[120] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996. The rights include "copyright, rights related to copyright or any other rights or obligations subsisting in the data, works or other materials incorporated into a database, patent rights, trade marks, design rights, the protection of national treasures, laws on restrictive practices and unfair competition, trade secrets, security, confidentiality, data protection and privacy, access to public documents, and the law of contract".

2.3.2 *Sui Generis Protection of Non-original Databases*

There is no obligation to protect non-original databases under Art. 10(2) of the TRIPS Agreement or Art. 5 of the WIPO Copyright Treaty. To date, there are no international norms on the protection of non-original databases.[121] This does not, however, imply that database rights are not valuable. Indeed, WIPO has acknowledged the importance of protecting databases for purposes of developing a global information infrastructure while at the same time ensuring the interests of users in having appropriate access.[122] The possibility of granting *sui generis* protection of databases that do not necessarily meet the threshold of originality in copyright law was introduced by the European Union during WIPO's diplomatic conference.[123] The proposal was, however, not pursued further at WIPO.[124] The United States also considered enacting laws to protect non-original databases from misappropriation but due to the ensuing controversies during the congressional debates no laws were enacted.[125] Consequently, only EU Member States grant *sui generis* protection for non-original databases.

Sui generis database protection was created by the EU Database Protection Directive to protect the "substantial investment in either the obtaining, verification or presentation of the contents to prevent extraction and/or re-utilization of the whole or of a substantial part".[126] This right is different from the copyright protection that the Directive grants for "databases which, by reason of the selection or arrangement of their contents, constitute the author's own *intellectual creation*".[127] This was emphasised in the case of *Football Dataco Ltd and Others v Yahoo! UK Ltd and Others*, where the European Court of Justice (ECJ) clarified that the purpose of the Database Directive is to "stimulate the creation of data storage and processing systems in order to contribute to the development of an information market … and not to protect the creation of materials capable of being collected in a database".[128] The ECJ further clarified that the requirement of the author's "own intellectual creation" for copyright protection[129] refers to the criterion of originality.[130]

[121] WIPO (2002), para. 7.

[122] WIPO (1996).

[123] WIPO, Basic Proposal for the Substantive Provisions of the Treaty on Intellectual Property in Respect of Databases Considered by the Diplomatic Conference on Certain Copyright and Neighbouring Rights Questions, Geneva CRNR/DC/6 (Dec. 1996).

[124] WIPO, Standing Committee on Copyright and Related Rights (SCCR/11/4, 16 September 2004).

[125] Davison (2016).

[126] Directive 96/9/EC, Art. 7(1).

[127] Directive 96/9/EC, Art. 3(1). The emphasis is added to show that the requirements for protection of the compilation of data are similar to those under Art. 10(2) of the TRIPS Agreement and Art. 5 of the WIPO Copyright Treaty.

[128] *Football Dataco Ltd and Others v Yahoo! UK Ltd and Others* (case C-604/10, 2012), para. 34.

[129] *See* Directive 96/9/EC, Art. 3(1).

[130] *Football Dataco Ltd and Others*, para. 37.

The scope of the *sui generis* right in Art. 7(1) of the Directive and the meaning of "a substantial investment in either the obtaining, verification or presentation of the contents" of the database were decided by the ECJ in *the British Horseracing Board* and *Fixtures Marketing* cases.[131] The Court stated that: "Article 7(1) of the directive reserves the protection of the *sui generis* right to databases which meet a specific criterion, namely to those which show that there has been qualitatively and/or quantitatively a substantial investment in the obtaining, verification or presentation of their contents."[132] This effectively excludes raw machine-generated databases and big data, which are typically drawn from multiple sources, from *sui generis* protection.[133] The ECJ also provided the following clarification regarding the expression "a substantial investment in either the obtaining, verification or presentation of the contents":

> [It has to] be understood to refer to the resources used to seek out existing independent materials and collect them in the database, and not to the resources used for the creation as such of independent materials. The purpose of the protection by the *sui generis* right provided for by the directive is to promote the establishment of storage and processing systems for existing information and not the creation of materials capable of being collected subsequently in a database.[134]

Consequently, the holder of the *sui generis* right can prohibit the manufacture of competing parasitical products and any actions that can cause significant detriment to the investment.[135]

The above clarifications evidently rule out any reliance on the sweat of the brow theory in granting *sui generis* rights in databases. As the European Commission correctly observed, *sui generis* rights are granted "to prevent misappropriation of the contents of a database in which there has been a substantial investment".[136] A distinction must therefore be made between "the establishment of storage and processing systems for existing information" and "the creation of materials capable of being collected subsequently in a database".[137] Investments in the establishment of the former are the object of *sui generis* rights, not the latter.[138] Evidently, *sui*

[131] *British Horseracing Board and Others v William Hill Organisation Ltd* (case C-302/02, 2004); *Fixtures Marketing Ltd v Svenska Spel AB* (case C-338/02, 2004); *Fixtures Marketing Ltd v Organismos prognostikon agonon podosfairou AE* (case C-444/02, 2004); *Fixtures Marketing Ltd v Oy Veikkaus Ab* (case C-46/02, 2004).

[132] *Fixtures Marketing Ltd v Oy Veikkaus Ab*, para. 32; *Fixtures Marketing Ltd v Svenska Spel AB*, para. 22.

[133] European Commission (2018), para. 5.4.1.

[134] *Fixtures Marketing Ltd v Oy Veikkaus Ab*, para. 34. *See also Fixtures Marketing Ltd v Svenska Spel AB*, para. 24.

[135] Directive 96/9/EC, Recital 42; *see also British Horseracing Board and Others*, para. 47.

[136] European Commission, written observations to the president and members of the Court of Justice of the European Union in *Football Dataco Ltd and Others v Yahoo! UK Ltd and Others* (case C-604/10) dated 15 April 2011, para. 23.

[137] *Fixtures Marketing Ltd v Organismos prognostikon agonon podosfairou AE*, para. 40.

[138] *See Fixtures Marketing Ltd v Svenska Spel AB*, paras. 31 and 33.

generis rights do not give rise to new rights in the works, data or materials that are contained in the databases.[139] Accordingly, *sui generis* rights should not be equated to IPRs that can be relied on to impede data sharing, since data as such are not owned by the party who incurs expenses on the investments. The investments are incurred to ensure the reliability of the information contained in the database, monitor the accuracy of the materials collected when creating the database and during its operation.[140] This essentially means that the scope of the investments is limited to the creation of the database.[141]

The effects of *sui generis* database rights on data sharing in health-related research should be considered in the context of a recent call, by UNESCO,[142] for data to be framed as a common good of humankind in line with Art. 2 of the Universal Declaration on Bioethics and Human Rights.[143] The Article requires the promotion of "equitable access to medical, scientific and technological developments as well as the greatest possible flow and the rapid sharing of knowledge concerning those developments and the sharing of benefits, with particular attention to the needs of developing countries".[144] *Sui generis* database rights may be used as a mechanism to impede the flow and rapid sharing of data that are contained in the protected database. This is the case because such rights are essentially used to control access to the data contained in the database such that any means of access that is considered to amount to misappropriation of the database is prohibited. This effect arises from the fact that *sui generis* database rights are modelled on laws that protect trade secrets or confidential information with a view to repressing any conduct that amounts to the "misappropriation" of an electronic database producer's investment.[145]

Notably, the European Commission has conceded that the wording that is used to describe the objectives of the Directive "suggests that the *sui generis* right may become a form of indirect property in data".[146] Right holders can rely on such proprietary claims over databases to restrict access to data for anti-competitive reasons, thus restricting data flows artificially.[147] Consequently, the emerging trend

[139] Directive 96/9/EC, Recital 46; *see also British Horseracing Board and Others*, para. 72.

[140] *British Horseracing Board and Others*, paras. 34 and 42; *Fixtures Marketing Ltd v Svenska Spel AB*, para. 27.

[141] *Fixtures Marketing Ltd v Svenska Spel AB*, para. 23.

[142] UNESCO (2017), para. 71. This call makes a lot of sense especially in the context of health research when it is considered from the perspective of the World Medical Association's Declaration of Taipei (2016), para. 5, which states that "health research represents a common good that is in the interest of individual patients, as well as the population and the society".

[143] UNESCO (2005).

[144] UNESCO (2005), para. f.

[145] Reichman and Samuelson (1997), p. 81. Trade secrets are further discussed in Sect. 2.3.3 of this paper.

[146] European Commission, (2018), para. 5.4.2.

[147] Duch-Brown et al. (2017). It is due to such effects that Reichman and Samuelson (1997), p. 88, predicted that under the Directive "every independent generation of data, however mundane or commonplace, will obtain protection if it costs money, and every regeneration or reutilization of the same data in updates, additions, and extensions that cost money will extend that protection without limit as to time".

of protecting non-original databases on the basis of substantial investment seems problematic due to over-protection in a research environment that is already facing challenges, particularly in LMICs where copyright protection of the database or the software that is required for the organisation, integration and analysis as well as production of data may require purchasing a licence, which is usually unaffordable for most LMICs.[148] Such protection essentially entails reliance on the substantial investment formula (protecting the value that is created in analysing the data),[149] which is admittedly very contentious in most jurisdictions since this leads to over-protection, thus restricting access to valuable information that is required for research and use by other interested stakeholders.[150] Additionally, such databases do not meet the threshold of originality, and factual information that is not original belongs to the intellectual commons, which should be accessed and used by interested stakeholders as appropriate.

A recent survey by the European Commission confirmed that *sui generis* rights have not achieved the intended purpose of incentivising the creation of databases; instead, they are mostly used in litigation when parties disagree.[151] Moreover, stakeholders from academia and research sectors indicated that the Directive did not achieve a balanced outcome in terms of safeguarding the legitimate interests of database makers and users.[152] The survey established that although there is no evidence that the *sui generis* regime itself leads to data lock-up, users found the licensing process complex due to additional layers of protection.[153] It also established that contract law is used to protect database owners' rights in addition to *sui generis* rights, thus leading academics and researchers to experience contractual overrides to the exceptions that are provided for in the Directive.[154] Clearly, creating *sui generis* rights for non-original databases was unnecessary.[155] Unfortunately, *sui generis* database rights are bound to continue existing and being used to unreasonably impede data sharing.

2.3.3 Trade Secrets

The purpose of protecting trade secrets is not to encourage secrecy or create any intellectual property rights. They merely protect the data against unfair misappropriation.[156] The law therefore provides for the enforcement of trust relationships in

[148] Luna et al. (2014), p. 38.

[149] Raju (2017), p. 219.

[150] Andanda (2016).

[151] European Commission (2018), para. 5.2.2.2.

[152] *Ibid*., para. 5.2.3.

[153] *Ibid*.

[154] *Ibid*., paras. 5.3.3.6 and 5.3.4.1; *see also* Malhotra (2016), an intellectual property law practitioner, who advises businesses to structure their contractual terms to specify that "any data derived from consumer transactions belong to the corporation". This shows how practitioners are encouraging companies to use contractual terms to create proprietary rights over data without any basis in intellectual property law, thus overriding the well-established legal principles that have been discussed in this paper.

[155] *See* Davison (2016), p. 852, who agrees with this view.

[156] Determann (2018).

this regard.[157] It is therefore not surprising that trade secret law has been labelled as "parasitic" because it relies on a host theory for normative support.[158] For example, it relies on other norms that are aimed at honouring contractual obligations and averting fraud for its existence.[159]

Data holders have relied on secrecy to protect their interests in data.[160] Such secrecy is based on Art. 39 of the TRIPS Agreement, which protects undisclosed information against unfair competition. Admittedly, this approach is rather controversial in protecting data in the context of big data since it is akin to erecting digital barbed wire around data that many deserving stakeholders are entitled to access. Reliance on trade secrecy in this regard is contestable since a proper interpretation of Art. 39 of the TRIPS Agreement confirms that the protection only extends to competition that is contrary to honest commercial practices, which is not the case for health-related research stakeholders.[161] The Article protects undisclosed information that has commercial value due to its being kept secret, thus ensuring business integrity. Consequently, loss of secrecy automatically leads to non-protection.

Secrecy or keeping information confidential to avoid sharing it with other stakeholders in health-related research goes against the intention of data subjects who consent to their information being used for research. As already noted, under the discussion of consent, commercialisation of data is not viewed favourably by data subjects. Secrecy erodes trust and can lead to data subjects declining to give their consent for the use of their data, thus stifling research and innovation.

The rationale of protecting trade secrets lies in the fact that the underlying information is generally unknown.[162] In the context of big data, data holders share very limited information on how data are collected (the factors considered) and the inferences drawn from the data.[163] This is mostly because methods of data preparation are viewed as valuable trade secrets, which have competitive advantage.[164] Withholding such vital information impedes the prospects of reusing, sharing and repurposing the data in a meaningful way. Although, as already clarified above, secrecy does not create IPRs, it defeats the purpose of protecting IPRs, which is to encourage sharing and dissemination of information.[165] This has led authors to wonder whether the IP regime should be amended to address the issue of non-disclosure in big data.[166] Clearly the solution lies in an alternative framework, as proposed in the next section, rather than amending the IP regime.

[157] Mittelstaedt and Mittelstaedt (1997).

[158] Bone (1998), p. 245.

[159] Bone (1998); Risch (2007).

[160] Malhotra (2016).

[161] *See* Andanda (2013) for an extensive discussion of the nature of related IP rights in health data and a proper interpretation of Art. 39 of the TRIPS Agreement.

[162] Risch (2007).

[163] Tene and Polonetsky (2013).

[164] Mattioli (2014) p. 549.

[165] Lemley (2008).

[166] Mattioli (2014).

Proponents of protection through trade secrecy have argued that "information-based processes that are not readily perceived by consumers are particularly well suited for trade secret protection".[167] This simply reinforces the argument that trade secrets are not IPRs since processes are not creative or inventive content that are capable of being protected through IP law.[168] Additionally, in health-related research, processes are important for follow-on research. Consequently, failure to disclose such information hinders further research and makes the generated data worthless (without the processes and insights that are drawn in the course of data analysis).

Other alternative mechanisms for incentivising data holders to share high quality data that may be the subject of considerable investments are already in use and should be considered instead of trade secrets. For example, medicine regulatory authorities may rely on the data for abridged approval of similar products submitted by competitors without disclosing the data.[169] Another option is ensuring that data holders' policies align with the FAIR data principles, namely making the data findable, accessible, interoperable and reusable.[170] A large network of international collaborators have developed FAIRsharing, which is an informative and educational resource that has adopted this approach for data management.[171]

3 Paradigm Shift from Ownership to Custodianship

The discussions in the preceding part of this paper have demonstrated that claiming ownership rights over data is misconceived because no such rights, over data as such, exist in the IP regime. It is clear that where individuals or companies claim to own data, such claims are either based on misinterpreting the scope of rights under copyright protection of compilations and *sui generis* rights over non-original databases or they use secrecy to avoid sharing data so that they can erroneously rely on Art. 39 of the TRIPS Agreement, or they use contractual terms to create proprietary rights that have no basis in the intellectual property regime. This confirms that ownership is a concept that is ill-suited for governing competing rights in big data. This fact finds fortification in the concerns that have been raised by authors such as Ekbia and colleagues that the law has not developed any "principle to balance the competing interests of individuals, industries, and society as a whole in the burgeoning age of Big Data".[172] The World Medical Association (WMA) has also urged relevant authorities to formulate policies and law that protect health data

[167] *Ibid*., p. 550.

[168] *See* the TRIPS Agreement, Art. 9(2), which excludes procedures and methods of operation from copyright protection, and Art. 27, which requires an inventive step for processes to be patentable. Abstract ideas such as algorithms that are used in big data are not protectable.

[169] For a detailed discussion of this alternative, *see* Andanda (2013), pp. 145–152.

[170] Sansone et al. (2019), p. 360.

[171] *Ibid.*

[172] Ekbia et al. (2015), p. 1535.

on the basis of the principles set forth in the Declaration of Taipei.[173] Custodianship is one of the principles of governance, which is stipulated in the Declaration.[174]

Due to claims of ownership over data, sharing and re-use of data may be restricted entirely or privileged access may be granted for a fee, or small data sets may be offered to university-based researchers.[175] Such practices deepen inequalities based on privileged access, mostly because data-owning companies have total control over data and no responsibility to make their data available, nor accountability to data subjects to ensure that their data are used in a manner that does not lead to harm. The ethical and governance challenges that beset Iceland in 1998 are very instructive in this regard. Serious issues arose from the declaration of health records, which included health, genetic and genealogical data, as a national resource that was owned by the Icelandic government and could be made available to private industry without the consent of the individuals.[176] As a result of national and international opposition to the inappropriate manner in which the Icelandic government handled the issue of ownership of data, the project collapsed in 2003.[177]

The IP regime is intended to stimulate creativity and not to protect non-proprietary matters such as underlying data or investments in creating databases as the current trend shows. As established in the previous part of the paper, data are not the object of monopoly rights under the IP regime since their generation is not the result of any creative endeavour. As the Hague Declaration on knowledge and discovery in the digital age succinctly puts it:[178] "Intellectual property was not designed to regulate the free flow of facts and ideas, but has as a key objective the promotion of research activity.… Licenses and contract terms should not restrict individuals from using facts, data and ideas."

This declaration is in line with Art. 9 of the TRIPS Agreement, which provides that copyright protection extends to expressions and not ideas. It essentially means that the urge to tap the full potential of big data must at the same time be accompanied by respect for other users' rights to access the information.

The legal framework in place mainly governs structured databases yet there is a massive amount of data that falls outside the scope of the current governance through ownership and IP. The reasons for this status quo are that current developments in big data have outpaced the existing legal framework[179] and big data practices do not fit within the frameworks of ownership and IP.[180] A paradigm shift from ownership to custodianship is warranted on two grounds: firstly, as already established in this paper, data are not the object of proprietary rights or ownership according to international IP law regime. Secondly, the emerging trends

[173] WMA (2016), para. 24.

[174] Article 20 requires custodians of health databases to consult and engage with individuals and their communities and to ensure accountability by being accessible and responsive to all stakeholders.

[175] Boyd and Crawford (2012), p. 673.

[176] Cook-Lucas et al. (2013); Winickoff (2006).

[177] Winickoff (2006).

[178] The Hague Declaration on Knowledge and Discovery in the Digital Age (2015), Principle 1.

[179] Tene and Polonetsky (2013), p. 241.

[180] Mattioli (2014).

that lead to claims of ownership over data are based on flawed models and on implausible arguments. The first point is extensively discussed in the preceding section of this paper. Therefore, this section focuses on advancing the second point.

One reason that is often advanced for claiming ownership rights over data is that a company or individual may have extracted new insights from original data, thus creating a new data set, which they should own. Sax observes that this argument is modelled on a "finders, keepers" ethic without due regard for the potential impact of the insights on the lives of data subjects.[181] The argument that the data in question may not be personal, thus warranting their appropriation and use without the data subject's consent may not be justifiable.[182] Additionally, as established under the discussion of copyright in compilations and *sui generis* database rights, investments in creating data are not taken into account in granting these types of proprietary rights in the IP regime. This confirms that there is no basis for claiming ownership rights in the newly created data set just because a company or individual has generated new insights from the original data.

Notably, all the five sources of big data that were highlighted in the introductory part of this paper are derived from personal information. Such information has accurately been described as expressing a sense of a person's "constitutive *belonging*, not of external *ownership*".[183] The criticism against viewing personal information through the lens of ownership further clarifies the constitutive nature of data such that it does not make sense to grant proprietary rights over it. The criticism, of relevance here, is the fact that one's personal information can never be lost when it is acquired by someone else.[184] In the context of big data, Sax has observed that "data that cannot be directly related to natural persons can be used, in big data contexts, to generate insights that can nonetheless have a significant impact on the lives and self-understanding of persons".[185] This observation is very instructive for appreciating that even the use of anonymised or de-identified data may be capable of re-identification due to the varying de-identification practices of data holders,[186] thus re-identifying the data subjects through the insights that are drawn from them.

Data subjects are entitled to informational privacy rights in their data. In essence, data subjects do not transfer their informational privacy rights to the parties who process their data. This approach can resolve the long-standing question of ownership of big data in health-related research. The position is that the data are not capable of being owned in the proprietary sense. The paradigm shift, which this paper advocates, entails recognising that researchers, commercial organisations and repositories that collect and process data have custodial rights and responsibilities in

[181] Sax (2016), p. 31.

[182] *See* the extensive discussion on the data subject's consent in this paper. *See also* Sax (2016), pp. 28–29 illustrating how the "finders, keepers" ethic can be applied in the context of ownership of big data.

[183] Floridi (2005), p. 195.

[184] *Ibid*., p. 194.

[185] Sax (2016), p. 30.

[186] Hoffman (2015), p. 1769.

handling the data.[187] The only proprietary rights that are capable of being owned are original compilations and not the data as such. As already explained, related rights are protected through copyright in the original compilation of the data or *sui generis* rights over non-original databases and trade secrets. It then becomes clear why arguments to the effect that "clinical trials data … are the property of the sponsoring company"[188] are neither accurate nor sustainable.

Granting property rights over underlying data is incapable of resolving the concerns in health-related research using big data, which have been discussed in this paper. New property rights in data will further impede data sharing, thus leading all stakeholders to lay claims over data. Such new rights can even lead to the emergence of data trolls who demand ransoms and nuisance fees based on potential property rights in data.[189] It thus makes sense to advocate a normative framework that is based on custodianship to ensure better accountability among stakeholders.

Custodianship has accurately been defined as "the responsibility for the safety and well-being of someone or something and represents ethical values like care, custody, … protection and trust to the guardianship or the safekeeping".[190] It is suitable for ensuring access to data and promoting fair data sharing practices while safeguarding data subjects' informational privacy at the same time. Appropriate custodianship of big data is necessary to ensure that data subjects maintain some control over access and future uses of their data while delegating decision-making in some matters to the data custodians. Such delegated decision-making gives rise to custodial rights, not ownership of the data.

So far, custodianship has been used as an ethical framework for ensuring shared accountability among all stakeholders involved in biospecimen-based research.[191] Although biospecimen-based research is significantly different from big data research, they have a common attribute in terms of claims of ownership that impede sharing of biospecimens or data, respectively. The salient feature of the framework is that it is based on ethical instead of "strictly legal principles to govern the collection and use of biospecimens in research".[192] As already established in the preceding discussions, reliance on the legal concept of ownership and even use of contractual terms have not resolved concerns in health-related research using big data. However, custodianship is much broader than the legal concept of ownership and has been used to ensure that all stakeholders recognise and honour their ethical obligations to serve the best interests of biomedical research.[193] The specific attributes of this proposed normative framework are discussed in the next part.

[187] *See* Mittelstadt and Floridi (2016), p. 327, who attribute custodial responsibilities to these categories of data controllers.

[188] *See* Vayena and Blasimme (2017), p. 505, who advance such an argument. *See also* Hoeren (2014), p. 754, who observes that "in general, the property in data is attributed to the originator, creator, or producer of these data".

[189] Determann (2018), p. 35.

[190] UNESCO (2017), para. 73.

[191] Yassin et al. (2010).

[192] *Ibid*., p. 1012.

[193] *Ibid.*

4 Attributes of the Alternative Normative Framework to Govern Rights in Big Data

The significance of health-related research lies in the fact that big data can be used for developing public health policies for disease surveillance and managing population health, hence the need for good governance. This requires properly governed access to data sets for research, including citizens' access to personal information, to avoid misunderstandings or bypassing doctor–patient relationships since medical professionals have to provide an accurate interpretation of the information in the data sets. Other valuable approaches that have been proposed for ensuring data sharing, such as data cooperatives,[194] would need the ethical framework of custodianship to function optimally especially after the cooperatives have granted access to research groups and other stakeholders.

UNESCO's recommendation that a framework with new approaches to ownership and custodianship of personal data be developed[195] as well as an appreciation of the fuzzy contours of data ownership and related IPRs are good starting points in highlighting two attributes of the alternative normative framework. Firstly, it makes a clear distinction between the underlying data and related IPRs in big data, thus ensuring respect for IPRs. Secondly, it is premised on ethical principles that can be used to manage diverse interests, thus effectively addressing concerns in health-related research using big data. These are explained below.

4.1 Distinguishing Between the Data and Related IPRs

Having established that there are components of big data that are protected through IPRs, it should be clear at this stage that the proposed normative framework should guide stakeholders in managing these IPRs in a manner that fosters data sharing. All stakeholders have custodial responsibilities over data since, even if they hold related IPRs, they need to fulfil custodial responsibilities over the underlying data that are not part of their monopoly rights. These responsibilities arise from their fiduciary relationship with the data subjects.[196] Consequently, it is essential to distinguish between the data and related IPRs.

The arguments that have been advanced in this paper essentially emphasise the fact that owners of related IPRs do not own the underlying data. They are custodians of the data, which should be made accessible to the interested stakeholders in accordance with the data subjects' consent. Citizens and data subjects are also key stakeholders in this regard, and they should be directly involved in the governance

[194] Blasimme et al. (2018), p. 475, have proposed data cooperatives as a way of ensuring that data subjects have direct control over their data and that they participate in the governance of the data. Such cooperatives would consist of individual data subjects who are "the most legitimate actors to promote personal data aggregation and to claim data control". The authors argue that clarity is needed on whether the prerogative of controllers relates to personal data ownership (p. 478). The clarity that is provided in this paper, on data ownership, and the proposed normative framework can be used for addressing this issue.

[195] UNESCO (2017).

[196] *See* Blasimme et al. (2018), p. 478, who also embrace the use of a fiduciary relationship in this context.

of big data.[197] Therefore, related IPRs should be managed to preserve open access to data and promotion of downstream commercialisation of inventions in a manner that fosters future research.[198]

4.2 Managing Interests in Data Through the Ethical Framework of Custodianship

This framework entails acknowledging data as a gift from data subjects to be used with their consent to advance science for the benefit of society and not to be owned by researchers, host institutions or funders/sponsors. The reason for this approach is that data subjects consider researchers who obtain their data to be custodians of the data. The custodial responsibilities entail compliance with rigorous ethical and regulatory requirements such as providing accurate and timely data and safeguarding data subjects' privacy and confidentiality.[199]

Using custodianship to govern rights in big data is premised on five ethical principles that are explained below:[200]

i. *Respect for privacy and autonomy*: This entails ensuring that measures are in place to protect data subjects' privacy while at the same time being accountable to them by maintaining open communication.[201]
ii. *Reciprocity*: Data custodians should provide feedback of the general results to relevant institutions and data subjects.[202] This principle can guide data controllers and other responsible stakeholders in ensuring the timely and efficient dissemination of aggregate research findings for the benefit of research participants and the public.[203]
iii. *Freedom of scientific enquiry*: This principle resonates with UNESCO's recommendation that stakeholders adopt an understanding of big data as a common good of humankind, hence the need to facilitate open access and use of data for the common good.[204]
iv. *Attribution*: As already noted, *sui generis* database rights can be used to restrict access to the underlying data. The principle of attribution can reduce such restrictive practices by ensuring that stakeholders acknowledge the substantial investments in creating the databases and mutually agree on the terms of use and access.
v. *Respect for intellectual property*: Although the underlying data are not the object of IP protection, stakeholders should respect related IPRs.

[197] Vayena et al. (2018).

[198] Yassin et al. (2010), p. 1014; Mascalzoni et al. (2015).

[199] Page et al. (2016); Mascalzoni et al. (2015).

[200] The principles are contained in the international charter of principles for sharing biospecimens and data; *see* details in Mascalzoni et al. (2015), p. 722.

[201] Ballantyne (2018).

[202] Mascalzoni et al. (2015).

[203] Yassin et al. (2010), p. 1014.

[204] UNESCO (2017).

If the above principles that are embedded in custodianship are applied, then the ideal framework that proponents of data ownership have suggested, namely one that can ensure better trust in the accuracy of the data and facilitate enhanced sharing,[205] can be established without granting ownership rights that risk reducing data to a commodity for profit, thereby restricting data sharing and reuse.

5 Conclusions

The key insights from the discussions in this paper are: firstly, data as such are not capable of being owned. However, this does not mean that they should not be protected through other mechanisms that are aimed at ensuring accountability instead of granting proprietary rights. Having established that the underlying data are not owned by anyone, it is safe to conclude that the IP regime does not need to adapt to current big data practices. What is urgently needed is an alternative normative framework that is based on the ethical principle of custodianship to ensure accountability and responsible data sharing by all stakeholders. Secondly, granting property rights over underlying data has no basis in IP law and is incapable of solving the concerns in health-related research using big data. An important lesson from the discussion of the scope of *sui generis* rights in databases, as discussed in this paper, is that the underlying raw data should not be protected as IPRs. Such property rights in data will further impede data sharing. The concerns can be addressed through the paradigm shift that is discussed in this paper, which entails recognising that researchers, commercial organisations and repositories that collect and process data have custodial rights and responsibilities in handling the data.

References

Alter GC, Vardigan M (2015) Addressing global data sharing challenges. J Empir Res Hum Res Ethics 10(3):317–323

Andanda P (2013) Managing intellectual property rights over clinical trial data to promote access and benefit sharing in public health. Int Rev Intellect Prop Compet Law 44(2):140–177

Andanda P (2016) Copyright law and online journalism: a South African perspective on fair use and reasonable media practice. Queen Mary J Intellect Prop 6(4):411–434

Ballantyne A (2018) Where is the human in the data? A guide to ethical data use. Gigascience 7(7):giy076. https://doi.org/10.1093/gigascience/giy076

Bellagio Big Data Workshop Participants (2014) Big data and positive social change in the developing world: a white paper for practitioners and researchers. https://www.rockefellerfoundation.org/report/big-data-and-positive-social-change-in-the-developing-world/. Accessed 26 Mar 2019

Bender E (2015) Big data in biomedicine: 4 big questions. Nature 527:S19

[205] *See* for example Kish and Topol (2015), p. 923.

Bettig VR (2018) Copyrighting culture: the political economy of intellectual property. Routledge, New York
Blasimme A, Vayena E, Hafen E (2018) Democratizing health research through data cooperatives. Philos Technol 31(3):473–479
Bone GR (1998) A new look at trade secret law: doctrine in search of justification. California Law Rev 86(2):241–313
Borgman LC (2012) The conundrum of sharing research data. J Am Soc Inf Sci Technol 63(6):1059–1078
Boyd D, Crawford K (2012) Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. Inf Commun Soc 15(5):662–679
Burtscher B, Fritz G (2015) Big data: who owns and who may use and exploit big data? https://www.lexology.com/library/detail.aspx?g=77ab3ffb-8f25-469c-ad80-d14bbcee9551. Accessed 7 Mar 2019
Cheung YSA (2018) Moving beyond consent for citizen science in big data health and medical research. Northwest J Technol Intellect Prop 16(1):15–40
Cook-Lucas JM, Schroeder D, Arnason G, Andanda P, Kimani J, Fournier V, Krishnamurthy M (2013) Donating human samples: who benefits? Cases from Iceland, Kenya and Indonesia. In: Schroeder D, Cook Lucas J (eds) Benefit sharing: from biodiversity to human genetics. Springer, Amsterdam, pp 95–127
Council of Europe, Directorate General of Human Rights and Rule of Law (2017) Guidelines on the protection of individuals with regard to the processing of personal data in a world of big data. T-PD(2017)01
Dalal P (2006) Data protection law in India: the TRIPS perspective. J Intellect Prop Rights 11:125–131
Davison M (2016) Database protection: lessons from Europe, Congress, and WIPO. Case West Res Law Rev 57(4):829–854
Denny GS, Silaigwana B, Wassenaar D, Bull S, Parker M (2015) Developing ethical practices for public health research data sharing in South Africa: the views and experiences from a diverse sample of research stakeholders. J Empir Res Hum Res Ethics 10(3):290–301
Determann L (2018) No one owns data. Hastings Law J 70(1):1–43
Dijcks J (2013) Oracle: big data for the enterprise. Oracle white paper. Oracle Corporation, Redwood Shores, CA
Duch-Brown N, Martens B, Mueller-Langer F (2017) The economics of ownership, access and trade in digital data. Digital economy working paper 2017-01. JRC technical reports
Ekbia H, Mattioli M, Kouper I, Arave G et al (2015) Big data, bigger dilemmas: a critical review. J Assoc Inf Sci Technol 66(8):1523–1545
Erikainen S, Pickersgill M, Cunningham-Burley S, Chan S (2019) Patienthood and participation in the digital era. Digit Health 5:205. https://doi.org/10.1177/2055207619845546
European Commission (2018) Commission staff working document: evaluation of Directive 96/9/EC on the legal protection of databases {swd (2018) 147 final}
European Commission, Directorate-General for Competition (2019). Competition policy for the digital era. A report by Crémer J, de Montjoye YA, Schweitzer H
European Commission, Directorate-General for Health and Consumers Unit D3 eHealth and Health Technology Assessment (2014) The use of big data in public health policy and research. Background information document. https://ec.europa.eu/health/sites/health/files/ehealth/docs/ev_20141118_co07b_en.pdf. Accessed 26 Sept 2019
European Commission, Directorate-General for Health and Food Safety, Directorate B – Health Systems, Medical Products and Innovation (2016) Study on big data in public health, telemedicine and healthcare. Final report. https://ec.europa.eu/health/sites/health/files/ehealth/docs/bigdata_report_en.pdf. Accessed 26 Sept 2019
Floridi L (2005) The ontological interpretation of informational privacy. Ethics Inf Technol 7(4):185–200
Hall AM (2010) Property, privacy, and the pursuit of interconnected electronic medical records. Iowa Law Rev 95:631–663
Hate K, Meherally S et al (2015) Sweat, skepticism, and uncharted territory: a qualitative study of opinions on data sharing among public health researchers and research participants in Mumbai, India. J Empir Res Hum Res Ethics 10(3):239–250
Haunss S, Shadlen K (2009) Introduction: rethinking the politics of intellectual property. In: Haunss S, Shadlen K (eds) Politics of intellectual property: contestation over the ownership, use, and control of knowledge and information. Edward Elgar Publishing, Cheltenham, pp 1–12
Hoeren T (2014) Big data and the ownership in data: recent developments in Europe. Eur Intellect Prop Rev 36(12):751–754
Hoffman S (2015) Citizen science: the law and ethics of public access to medical big data. Berkeley Technol Law J 30(3):1741–1806

Innes P (2010) Ethical problems in archival research: beyond accessibility. Lang Commun 30:198–203
International Telecommunication Union (2016) ICT facts and figures 2016, ITU
Kallinikos J, Tempini N (2014) Patient data as medical facts: social media practices as a foundation for medical knowledge creation. Inf Syst Res 25(4):817–833
Kaye J, Whitley AE, Lund D, Morrison M, Teare H, Melham K (2015) Dynamic consent: a patient interface for twenty-first century research networks. Eur J Hum Genet 23:141–146
Kish JL, Topol JE (2015) Unpatients – why patients should own their medical data. Nat Biotechnol 33(9):921–924
Laney D (2001) 3D data management: controlling data volume, velocity, and variety. Technical report, META Group
Lemley AM (2008) The surprising virtues of treating trade secrets as IP rights. Stanf Law Rev 61(2):311–353
Lipworth W, Mason HP, Kerridge I, Ioannidis APJ (2017) Ethics and epistemology in big data research. Bioethical Inq 14:489–500
Luna DR, Mayan JC, García MJ, Almerares AA, Househ M (2014) Challenges and potential solutions for big data implementations in developing countries. IMIA Yearb Med Inform 23:36–41
Malhotra P (2016) How big data and IP intersect: big data is big business – but who owns it? Intellect Prop, an ALM Supplement (Fall)
Mascalzoni D, Dove SE, Rubinstein Y et al (2015) International charter of principles for sharing biospecimens and data. Eur J Hum Genet 23:721–728
Mattioli M (2014) Disclosing big data. Minnesota Law Rev 99(2):535–585
Megget K (2011) Riding the data stream. PharmaTimes Digital. http://www.pharmatimes.com/Magazine/Riding_the_data_stream.aspx. Accessed 4 Oct 2017
Mellado B (2015) The big data challenge and how Africa can benefit. The conversation, 19 November 2015
Merson L, Phong TV, Nhan LNT, Dung NT, Ngan TTD, Kinh NV (2015) Trust, respect, and reciprocity: informing culturally appropriate data-sharing practice in Vietnam. J Empir Res Hum Res Ethics 10(3):251–263
Mittelstadt DB, Floridi L (2016) The ethics of big data: current and foreseeable issues in biomedical contexts. Sci Eng Ethics 22:303–341
Mittelstaedt DJ, Mittelstaedt AR (1997) The protection of intellectual property: issues of origination and ownership. J Public Policy Mark 16(1):14–25
Mogha P, Sharma N, Sharma S (2013) Big data. Int J Res Inf Technol 1(11):223–230
Nicen P (2015) Better insights, better drugs. Nature 527:S18
OECD (2017) OECD recommendation on health data governance. Paris, OECD. http://www.oecd.org/els/health-systems/health-data-governance.htm. Accessed 26 Mar 2019
Ostherr K, Borodina S, Bracken RC, Lotterman C, Storer E, Williams B (2017) Trust and privacy in the context of user-generated health data. Big Data Soc. https://doi.org/10.1177/2053951717704673
Page SA, Manhas KP, Muruve DA (2016) A survey of patient perspectives on the research use of health information and biospecimens. BMC Med Ethics 17(1):48. https://doi.org/10.1186/s12910-016-0130-4
Parker M, Bull S (2015) Sharing public health research data: toward the development of ethical data-sharing practice in low- and middle-income settings. J Empir Res Hum Res Ethics 10(3):217–224
Pentland A, Reid GT, Heibeck T (2013) Revolutionizing medicine and public health: report of the big data and health working group. World Innovation Summit for Health (WISH)
Raju DK (2017) Database protection in India: need for reforms. In: Sinha MK, Mahalwar V (eds) Copyright law in the digital world: challenges and opportunities. Springer, Berlin, pp 205–220
Reichman HJ, Samuelson P (1997) Intellectual property rights in data? Vanderbilt Law Rev 50:51–166
Reinbothe J, von Lewinski S (2015) The WIPO treaties on copyright: a commentary on the WCT, the WPPT, and the BTAP. Oxford University Press, Oxford
Risch M (2007) Why do we have trade secrets? Marquette Intellect Prop Law Rev 11(1):1–76
Ritter J, Mayer A (2018) Regulating data as property: a new construct for moving forward. Duke Law & Technol Rev 16(1):220–277
Rumbold MMJ, Pierscionek B (2017) The effect of the general data protection regulation on medical research. J Med Internet Res 19(2):e47
Sack DA, Brooks V, Behan M et al (2009) Improving international research contracting. Bull World Health Organ 87:487–487A
Sankor O, Ijsselmuiden C (2011) Sharing research data to improve public health: a perspective from the global south. Lancet 378:401–402

Sansone SA, McQuilton P, Rocca-Serra P et al (2019) FAIRsharing as a community approach to standards, repositories and policies. Nat Biotechnol 37:358–367
Sax M (2016) Big data: finders keepers, losers weepers? Ethics Inf Technol 18(1):25–31
Shah N, Coathup V, Teare H et al (2018) Sharing data for future research – engaging participants' views about data governance beyond the original project: a DIRECT study. Genet Med. https://doi.org/10.1038/s41436-018-0299-7
Steinsbekk KS, Myskja KB, Solberg B (2013) Broad consent versus dynamic consent in biobank research: is passive participation an ethical problem? Eur J Hum Genet 21:897–900
Tene O, Polonetsky J (2013) Big data for all: privacy and user control in the age of analytics. Northwest J Technol Intellect Prop 11(5):239–273
The Berne Convention for the Protection of Literary and Artistic Works (1886)
The Hague Declaration on Knowledge and Discovery in the Digital Age (2015) https://thehaguedeclaration.com/the-hague-declaration-on-knowledge-discovery-in-the-digital-age/. Accessed 26 Mar 2019
The IET and Royal Academy of Engineering (2015) Response to the big data dilemma inquiry. https://www.raeng.org.uk/publications/responses/big-data-dilemma. Accessed 26 Sept 2019
Thorpe JH, Gray EA (2015) Big data and public health: navigating privacy laws to maximize potential. Public Health Rep 130(2):171–175
UNESCO (2005) Universal Declaration on Bioethics and Human Rights. Paris, UNESCO. http://portal.unesco.org/en/ev.php-URL_ID=31058&URL_DO=DO_TOPIC&URL_SECTION=201.html. Accessed 8 Dec 2018
UNESCO (2015) Report of the International Bioethics Committee of UNESCO (IBC) on the principle of the sharing of benefits. Paris, UNESCO. http://unesdoc.unesco.org/images/0023/002332/233230E.pdf. Accessed 10 Dec 2018
UNESCO (2017) Report of the International Bioethics Committee of UNESCO (IBC) on big data and health. Paris, UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000248724. Accessed 10 Dec 2018
van Panhuis GW, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst JA, Heymann D, Burke SD (2014) A systematic review of barriers to data sharing in public health. BMC Public Health 14:1144
Vayena E, Blasimme A (2017) Biomedical big data: new models of control over access, use and governance. Bioethical Inq 14:501–513
Vayena E, Dzenowagis J, Brownstein SJ, Sheikh A (2018) Policy implications of big data in the health sector. Bull World Health Organ 96:66–68
Ward JS, Barker A (2013) Undefined by data: a survey of big data definitions. Cornell University Library, Ithaca
Winickoff DE (2006) Genome and nation: Iceland's health sector database and its legacy. Innovations 1(2):80–105
WIPO (1996) Diplomatic conference on certain copyright and neighbouring rights questions, Geneva, 2–20 December 1996. In: Recommendation concerning databases, adopted by the diplomatic conference on 20 December 1996
WIPO (2002) Standing committee on copyright and related rights, Eighth Session, Geneva, 4–8 November 2002. Summary on existing legislation concerning intellectual property in non-original databases
World Health Organisation (2016) Global diffusion of eHealth: making universal health coverage achievable. The third global survey on eHealth. http://apps.who.int/iris/bitstream/10665/252529/1/9789241511780-eng.pdf. Accessed 26 Mar 2019
World Medical Association (2016) Declaration of Taipei on ethical considerations regarding health databases and biobanks. https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks/. Accessed 26 Sept 2019
Wyber R, Vaillancourt S, Perry W, Mannava P, Folaranmi T, Leo Celi LA (2015) Big data in global health: improving health in low- and middle-income countries. Bull World Health Organ 93:203–208
Yassin R, Lockhart N, González del Riego M et al (2010) Custodianship as an ethical framework for biospecimen-based research. Cancer Epidemiol Biomark Prev 19(4):1012–1015
Zwitter A (2014) Big data ethics. Big Data Soc 1(2):2053951714559253