

# Big Tech platforms in health research: Re-purposing big data governance in light of the General Data Protection Regulation's research exemption

Big Data & Society  
January–June: 1–14  
© The Author(s) 2021  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20539517211018783  
journals.sagepub.com/home/bds  
 SAGE

Luca Marelli<sup>1,2,3</sup> , Giuseppe Testa<sup>3,4,5</sup> and  
Ine Van Hoyweghen<sup>1,6</sup>

## Abstract

The emergence of a global industry of digital health platforms operated by Big Tech corporations, and its growing entanglements with academic and pharmaceutical research networks, raise pressing questions on the capacity of current data governance models, regulatory and legal frameworks to safeguard the sustainability of the health research ecosystem. In this article, we direct our attention toward the challenges faced by the European General Data Protection Regulation in regulating the potentially disruptive engagement of Big Tech platforms in health research. The General Data Protection Regulation upholds a rather flexible regime for scientific research through a number of derogations to otherwise stricter data protection requirements, while providing a very broad interpretation of the notion of “scientific research”. Precisely the breadth of these exemptions combined with the ample scope of this notion could provide unintended leeway to the health data processing activities of Big Tech platforms, which have not been immune from carrying out privacy-infringing and socially disruptive practices in the health domain. We thus discuss further finer-grained demarcations to be traced within the broadly construed notion of scientific research, geared to implementing use-based data governance frameworks that distinguish health research activities that should benefit from a facilitated data protection regime from those that should not. We conclude that a “re-purposing” of big data governance approaches in health research is needed if European nations are to promote research activities within a framework of high safeguards for both individual citizens and society.

## Keywords

General Data Protection Regulation, research exemption, Big Tech platforms, data governance

## Introduction

In November 2019, the Wall Street Journal broke the news that the second-largest healthcare provider in the US, run by the non-profit organization Ascension, had entered a deal with Google that could eventually see detailed health information on 50 million American patients moving into the company's cloud-computing system (Copeland and Needleman, 2019). The massive data transfer is part of an agreement—code name: “Project Nightingale”—whose stated aim has been to employ Google Cloud's machine learning tools for developing new diagnostic tests and technologies for exchange of patient health information (Cohen, 2019). At the outbreak of the news, the two parties involved were swift in publicly claiming no legal wrongdoing and allay fears over repurposing of patient data,

such as for targeted advertising (Conrado, 2019; Shaukat, 2019). Nonetheless, the news of Ascension's

<sup>1</sup>Life Sciences & Society Lab, Centre for Sociological Research, KU Leuven, Leuven, Belgium

<sup>2</sup>Department of Medical Biotechnologies and Translational Medicine, University of Milan, Milan, Italy

<sup>3</sup>Department of Experimental Oncology, IEO, European Institute of Oncology IRCCS, Milan, Italy

<sup>4</sup>Department of Oncology and Hematology-Oncology, University of Milan, Milan, Italy

<sup>5</sup>Human Technopole, Milan, Italy

<sup>6</sup>KU Leuven Institute for Single Cell Omics (LISCO), Leuven, Belgium

### Corresponding author:

Luca Marelli, KU Leuven, Parkstraat 45, Leuven 3000, Belgium.  
Email: [luca.marelli@kuleuven.be](mailto:luca.marelli@kuleuven.be), [luca.marelli@unimi.it](mailto:luca.marelli@unimi.it),  
[luca.marelli@ieo.it](mailto:luca.marelli@ieo.it)



sharing with Google of identifiable information on millions of patients, without notification to either patients or doctors, sparked an intense public outcry, in the US and also beyond (Copeland and Needleman, 2019; Prainsack, 2020).

Whilst much scrutinized, Google's deal with Ascension represents but one of the latest examples of an increasing trend that, over the space of a decade, has seen the health sector being flooded by a growing stream of digital technology corporations—Google, but also Amazon, Microsoft, IBM, and Apple, to name but the largest—jockeying for position in this rapidly expanding market (Sharon, 2016). Google itself, in response to the scandal, has been plain in acknowledging that similar agreements have been established “with dozens of other healthcare providers” (Shaukat, 2019). In fact, the company is said to have gained access to tens of millions of patient health records in more than three-quarters of US states, consisting in many instances of personally identifiable health information (Copeland et al., 2020). And Google is not alone in this endeavor. All its competitors have moved equally aggressively to claim their stakes in the ever-expanding field of digital health.

Not surprisingly, the emergence of a global industry of digital health platforms<sup>1</sup> operated by Big Tech corporations has raised pressing questions on the capacity of current data governance models, regulatory and legal frameworks to safeguard the sustainability of the digital health ecosystem, through the promotion of mutually beneficial interactions between digital health platforms and society (Taylor and Purtova, 2019; Van Dijck et al., 2018).

In this article, we direct our attention toward the data governance regime unfolding in the European Union (EU) around Regulation (EU) 2016/679, the General Data Protection Regulation (GDPR), which became applicable since 25 May 2018. Establishing a comprehensive and harmonized data protection framework across the EU, the GDPR has been lauded in many quarters as the cornerstone of Europe's value-friendly “third way” to navigate the digital world, in between the “techno-libertarianism” of US Silicon Valley and the “digital authoritarianism” of China (Thornhill, 2018). Yet, as we reviewed in recent work (Marelli et al., 2020), the GDPR faces mounting challenges in regulating the entanglements of digital health technologies and related big data practices that, as means for “intensified data sourcing” (Hoeyer, 2016: 74), are increasingly pervading our digitized European societies.

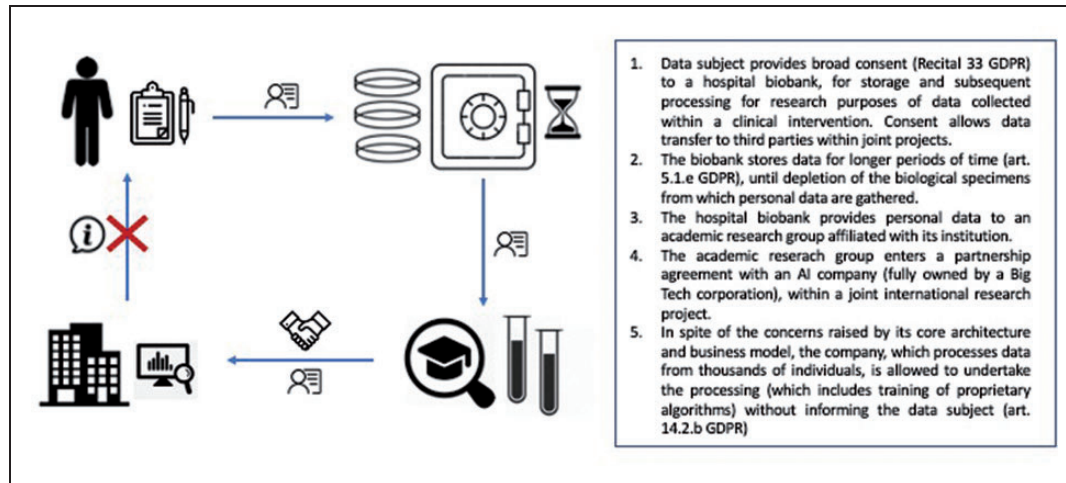
Within this scenario, this article specifically focuses on the GDPR's provisions that directly target health research. Notwithstanding its “omnibus” rather than sectoral approach to data protection, intended to cover a wide scope of processing areas, the GDPR

devotes a specific set of provisions to regulate the processing of personal data for scientific research activities. Notably, the GDPR ushers in a rather permissive and flexible regime for scientific research through a number of derogations to otherwise stricter data protection requirements, while providing a very broad interpretation of the notion of “scientific research”. In fact, we contend, the breadth of these derogations combined with the ample scope of this notion can provide unintended leeway to the health data processing activities of Big Tech platforms, which have not been immune from carrying out privacy-infringing and socially disruptive practices in the health domain. Accordingly, it is the contention of this article, further finer-grained demarcations need to be traced in order to distinguish research activities that can legitimately avail themselves of data protection facilitations from those that mandate for stricter data protection standards.

The article is structured as follows. First, we reconstruct the policy trajectory that inscribed facilitations for scientific research into the GDPR. Next, we outline the bundle of provisions that make up the so-called “research exemption” and critically discuss the notion of scientific research in the GDPR. Then, we review the challenges and potential disruptions elicited by the influence of Big Tech platforms in the health domain, partly owing to the data protection facilitations originally intended to streamline and de-bureaucratize established and well-regulated research practices. Finally, we provide an overview of and discuss proposals for deploying use-based data governance frameworks to demarcate research activities to be carried out within a facilitated data protection regime from those to subject to more stringent data protection scrutiny. This discussion is intended as the basis for the further development of more detailed policy guidance—such as the one presently in elaboration at both national and European level (European Data Protection Board (EDPB), 2021).

## **Making the GDPR into law: Policy struggles toward a special derogatory regime for scientific research**

The approval of the GDPR has been characterized by a lengthy and contentious policy process. “Trilogue” negotiations amongst the European Commission, the European Parliament, and the Council of Europe<sup>2</sup> were shaped by intense lobbying efforts, mostly by digital corporations, and brought to the surface underlying lines of tension among EU Member States diverging in economic interests and cultural understandings on privacy and data protection (Mager, 2017). Throughout the process, almost 4000



**Figure 1.** A stylized representation of how the research exemption could apply in practice.

amendments had been tabled, making the GDPR the most lobbied legislation in EU history.

A number of actors and organizations associated with (data-intensive) health research (from the Wellcome Trust to Science Europe) were prominently involved in the negotiating phases, as they set out to shift the policy discourse from data- to health-protection through sustained generation and application of scientific knowledge (Starkbaum and Felt, 2019). In so doing, they took aim at drafts of the legislation, most notably the one approved in March 2014 by the Committee on Civil Liberties, Justice and Home Affairs of the European Parliament, deemed to adopt a too restrictive approach to regulating research (Casali, 2014; Ploem et al., 2013). Overall, the aim of these converging lobbying efforts from the part of individual researchers and research agencies was to ensure “that the needs of scientific research with respect to accessibility and processing of personal data are considered, that the provisions and derogations that facilitate scientific research are maintained, and that amendments which would dramatically weaken these provisions are rejected” (Science Europe, 2013: 2).

A settlement in Trilogue negotiations was reached in December 2015, and, notwithstanding some critical remarks (e.g. LERU, 2016), the definitive text of the GDPR was met with a general sense of escaped danger from the part of the European health research community (Abbott, 2015; Dove et al., 2016). For, the outcome of these negotiations resulted in the GDPR arguably providing a significant degree of flexibility in the use of personal data for scientific research purposes, stemming from the implicit recognition of the *sui generis* status of research as a social activity warranting tailor-made provisions and derogations from the most stringent data protection requirements.

## The research exemption in the GDPR

The “special derogatory regime” (EDPB, 2021) for scientific research in the GDPR—also known as the “research exemption”—contains four distinct sets of provisions. The first, set out in Art. 9(2)(j), provides an alternative to the explicit consent requirement for personal data processing outlined in Art. 9(2)(a). Pursuant to Art. 9(2)(j), scientific research is identified as one of the processing activities for which the general prohibition otherwise imposed in Art. 9(1) on the processing of “special categories of personal data” (i.e. sensitive data, such as health-related and genetic data) is withdrawn; what is more, scientific research is identified as one of the lawful grounds for processing sensitive data, even in absence of consent from the data subject (Art. 9(2)(j) read in conjunction with Art. 6(1)(f)).<sup>3</sup> To ensure the lawfulness of the processing, scientific research activities should meet two additional conditions. First, they should be subject to appropriate organizational and technical safeguards for data subjects, as set out in Art. 89(1), in particular to ensure the principle of data minimization (which limits the use of data to what is necessary to achieve the stated purpose of the processing). In addition, the research should be undertaken on the basis of a specific national or EU law providing a relevant legal basis. Such law should be proportionate to the aim pursued, respect the essence of the right to data protection, and must be interpreted in consideration of the jurisprudence of the European Court of Justice (ECJ) (EDPB, 2020).

Second, meeting the demands of research organizations, the Regulation introduces much called-for provisions geared to facilitate the re-use of data in secondary research. The GDPR provides that secondary processing of previously collected data for scientific research purposes is “not incompatible” with the initial



purpose for which data has been gathered (Art. 5(1)(b)), thus making possible to re-use data without consent from the data subject. This “presumption of compatibility” (EDPS, 2020: 22)—however—does not provide a blanket authorization to secondary research. Rather, as recently observed by the “*Preliminary opinion on data protection and scientific research*” of the European Data Protection Supervisor (EDPS), “each case must be considered on its own merit and circumstances” (EDPS, 2020: 22; cf. also Article 29 Working Party, 2013: 28), in line with the contextual criteria for compatibility assessment delineated in Art. 6(4). Still, this provision establishes an important principle bound to facilitate the re-use of data for secondary research.

Additionally, in those cases in which consent is used as a legal basis, the GDPR allows data controllers (the entities legally in control of data processing) to avail themselves of “broad consent”, whenever this is required by the intended research purposes, and especially when it is not possible to fully identify the purposes of personal data processing at the time of data collection, as eminently in the case of biobanking (recital 33; cf. also EDPB, 2021; Hallinan, 2020; Marelli and Testa, 2018; Shabani et al., 2021). In such instances, data controllers are required to provide adequate safeguards to data subjects and describe the purposes of data processing in “a high-level way, for instance in terms of (types of) research questions and/or fields of research to be explored” (EDPB, 2021: 7).

The GDPR also relaxes limitations in the amount of time that data can be stored (Art. 5(1)(e)), and recognizes the importance to take into account “the legitimate expectations of society for an increase of knowledge” (Recital 113) where assessing whether the legitimate interest of the data controller can be invoked for (limited) cross-border data transfers outside of the EU.<sup>4</sup>

Third, the GDPR provides for a number of derogations to data subjects’ rights where the exercise of such rights is likely to render impossible or seriously impair the accomplishment of research objectives.<sup>5</sup> A first subset of derogations to data subjects’ rights is directly inscribed in the GDPR, which withholds the exercise of the “right to be forgotten” (which allows data subjects to have their data erased), and the right to receive information from data controllers which have not themselves collected the data, as it could be the case with data retrieved from patients’ records (Art. 17(3)(d) and Art. 14(5)(b)). Pursuant to Art. 89(2), a second subset of derogations related to scientific research undertakings—namely to the rights of access (Art. 15), rectification (Art. 16), restriction of processing (Art. 18), and objection (Art. 21)—can instead be introduced by Member States in their national legislations. However, as noted by the EDPB, these provisions

should be interpreted in the light of the jurisprudence of the ECJ, meaning that all restrictions of the rights of data subjects must apply only insofar as strictly necessary (EDPB, 2020: 11).

Finally, in a move that has been regarded as potentially leading to the further fragmentation of the European health research landscape (LERU, 2016), Member States are also endowed with the prerogative to introduce further provisions for the processing of genetic, health, and biometric data (Art. 9(4)). Whilst these could also include limitations to research, countries such as Italy have been prompt in passing legislation that can be seen as highly favorable for health research, for instance by providing that sensitive data collected for clinical purposes by Italian research hospitals—the 51 *Istituti di Ricovero e Cura a Carattere Scientifico* certified by the Health Ministry—can be re-purposed for scientific research without the need for additional consent given that the medical care they provide is said to be “instrumental” to their scientific research activities (Decreto Legislativo 10 agosto 2018, n. 101, Art. 110-bis).

## The notion of “scientific research” in the GDPR

In defining the scope of the special derogatory regime for scientific research, the legislator has apparently confronted a contemporary uptake of a foundational epistemic problem—what should qualify as genuine science? In that regard, the GDPR advances a remarkably broad interpretation of activities falling within the remit of what should count as scientific research proper. As stated in Recital 159:

*For the purposes of this Regulation, the processing of personal data for scientific research purposes should be interpreted in a broad manner including for example technological development and demonstration, fundamental research, applied research and privately funded research. In addition, it should take into account the Union’s objective under Article 179(1) TFEU of achieving a European Research Area.*

On the one hand, this interpretation—which, it should be noted, falls short of representing either a *formal* or a *substantive normative* definition of scientific research (Buttarelli, 2018; Van Veen, 2018)—reflects the inevitable predicament of defining research while having to account for its increasingly heterogenous landscape (especially in contemporary health research) (Carrier and Nordmann, 2011; Nowothny et al., 2001). On the other hand, it reflects the GDPR’s “omnibus”, rather than sectoral, approach to data protection, which

enables the Regulation to cover a wide scope of processing areas and different types of processing activities. At the same time, it is informed by EU science policy priorities of constructing a European knowledge society, by integrating public and private resources and know-hows (Marelli and Testa, 2017), and by “opening up the innovation process to people with experience in fields other than academia and science” (European Commission, 2016).

At any rate, such a broad construe extends the scope and reach of the special derogatory regime for research. Also depending on additional provisions concerning scientific research enacted at the national level (for a comprehensive review, see European Commission, 2021), the bundle of facilitations traced above stands to apply not only to public and private academic or clinical research centers, but also to actors such as pharmaceutical, medical device, and—as per the focus of this article—digital technologies companies undertaking health research activities.

### **How broadly should the GDPR’s notion of scientific research be interpreted?**

In the view of the EU legislator, public officials, and data protection bodies, “the definition of scientific research must not be stretched beyond its intended limits” (Buttarelli, 2018: 2), and processing operations linked to scientific research must “respect the essence of the right to data protection” (Art. 9(2)(j) GDPR). Likewise, as claimed by the EDPS, the EU’s independent data protection supervisory authority, “performing an activity deemed to be research cannot be a *carte blanche* to take irresponsible risks. From a data protection viewpoint, the principles of necessity and proportionality are essential” (EDPS, 2020: 11). Moreover, as remarked by the Article 29 Working Party, “scientific research” should be understood as “a research project set up in accordance with relevant sector-related methodological and ethical standards, in conformity with good practice” (Article 29 Working Party, 2017: 27). Thus, as summarized by the EDPS, for the special data protection regime to apply, research should be performed in accordance with ethical standards, and with the aim of “growing society’s collective knowledge and wellbeing, as opposed to serving primarily one or several private interests” (EDPS, 2020: 12).

And yet the above points of caution appear, in a number of respects, of difficult implementation. For one thing, it is questionable whether—and especially *how*—core principles of privacy and data protection, like the ones of necessity and proportionality recalled by the EDPS, can apply to current developments in the

field of digital health research. As we reviewed more in depth elsewhere (Marelli et al., 2020), these and other key principles (e.g. purpose limitation, data minimization, transparency) are not only technically challenged but, epistemically, on an overt collision course with the very thrust of technologies and research practices that are meant, by design, to enable extensive and indeed open-ended data processing and repurposing of previously collected data (e.g. artificial intelligence (AI)). In turn, the open-endedness of such iterative repurposing also grounds much of the business models embraced by the novel types of actors engaged in research (notably Big Tech platforms). This tension, in turn, could make it possible that either the provisions set forth by the GDPR are rendered “quickly irrelevant” (Zarsky, 2017: 996), thus lowering the level of protection afforded to citizens in Europe, or that their stringent enforcement ends up stifling the potential of European digital innovation toward suboptimal and globally marginalizing approaches (Chivot, 2019).

It is also questionable whether current ethical standards and modes of ethical oversight provide sufficient bite and are adequately tailored for governing current developments in health research. For instance, as preliminary data shows,<sup>6</sup> research ethics committees—which have long since played a central role for upholding ethical standards in research—have to confront the lack of well-established guidelines and regulatory tools for providing adequate ethics assessment for non-traditional types of data-intensive research (e.g. AI, social companion robots). In parallel, inasmuch as they have been traditionally appointed with the aim to review clinical research on human subjects, they are in many cases devoid of the epistemic as well as ethical expertise required to master recently established research fields and deal with the ethical quandaries raised by new data-intensive technologies and research practices. This, as a consequence, can lower standards of ethical oversight in health research, to the potential detriment of research participants (Staunton et al., 2019).

Likewise, the reference to “sector-related methodological standards” seems a rather ineffective if not counter-productive move to avoid over-stretching the limits of what should count as scientific research. In contemporary health research, which increasingly aligns heterogeneous technologies and expertise, what is regarded as sound “methodological standards” is the result of ongoing and open-ended processes of coordination and convention-making (Cambrosio et al., 2006; Keating and Cambrosio, 2003). As the latter involve a multiplicity of actors spanning distinct sectors, the very notion of “scientific sector” appears something increasingly challenged and in flux. Moreover, in the case of digital health, a sizeable portion of actors that play a

central role in standard-setting (e.g. digital technology companies) are primarily involved in activities *other* than scientific research. The latter is thus potentially shaped by, and increasingly made porous to, methodologies and modes of practice that are not necessarily aligned with epistemic standards, but also norms and institutional values, traditionally associated with specific scientific domains (see e.g. discussions around the opacity of algorithms, or the lack of reproducibility for AI systems in health) (Burrell, 2016; McDermott et al., 2019). Consequently, if the recourse to sectoral methodological standards to define scientific research had to be taken at face value, this would unduly narrow the scope of health research to a limited number of practices—to the paradoxical extent that the most innovative research activities would be placed outside the remit of the definition of health research.

As to the remarks about the overall primary orientation of research toward “collective knowledge and wellbeing” (rather than private interests), the difficulty here is twofold. First, consistently with what we have just observed, scientific research in the health domain has increasingly come to be carried out in a “transdisciplinary” manner in so-called “contexts of application”, where it is difficult to neatly discriminate between undertakings that primarily pursue collective knowledge and wellbeing as opposed to private interests (Carrier and Nordmann, 2011; Nowothny et al., 2001; Verschraegen, 2018). Second, from a technical legal point, it should be observed that the notion of “collective knowledge and wellbeing” is arguably subsumed in the GDPR under the notion of “public interest”. The scope of this notion, which can apply for instance to epidemiological studies (Article 29 Working Party, 2014), has to be defined in specific and precise enough terms in European and national legislation (Art. 6(3)(a)/(b) GDPR, Article 29 Working Party, 2014), and thus maintains a narrower connotation with respect to what, as a society, we would normally be prepared to identify, in lay terms, as the “general interest” or the “public good”. From the same legal perspective, it should be noted that “public interest” represents but *one* among the several legal bases, along e.g. the legitimate interest of data controllers, which can be invoked for providing legitimate justification for data processing in scientific research without consent from data subjects. From this, it follows that scientific research need not be pursued on the basis of a public interest, in order to avail itself of the facilitation mechanisms introduced by the GDPR.

As a consequence of these two points, in absence of a better definition of what the pursuit of collective knowledge and wellbeing should amount to, the

observation that research should be geared to some forms of public good appears of difficult actionability.

As it stands, and notwithstanding claims to the contrary by EU officials and data protection bodies, the notion of scientific research advanced in the GDPR appears exceedingly broad-construed. Consequently, questions remain as to whether the ad hoc regime for scientific research maintains an excessive reach, in ways perhaps unintended by the legislator.

The relaxation of data protection requirements was foreseen to streamline and de-bureaucratize research practices traditionally subject to high ethical and regulatory standards, in view of the potential societal benefits accruing from (health) research. Still, depending on how Member States modulate the research exemption in their own legislation (see European Commission, 2021), and considering especially fields of research involving the processing of highly sensitive and identifiable data such as genomics (Shabani and Marelli, 2019), it can be argued that the GDPR risks reaching a paradoxical outcome. While wanting to carve out a special status for scientific research (thereby implicitly stating that not every data processing activity can fall under its rubric and that the sheer pervasiveness of the digital has not yet eroded its contours), it has broadened its definition so much that it may end up enabling, in practice, the erosion of the very status that it strived to protect. This may possibly lead to a major regulatory leeway in favor of data controllers over data subjects (Pormeister, 2017), reducing the accountability of research organizations, while depriving research participants of decisional autonomy and adequate safeguards.

Whilst the argument about the GDPR providing excessive regulatory leeway could be plausibly applied to the processing activities of different types of data controllers, including pharmaceutical corporations (Marelli and Testa, 2018), the reach of such leeway raises concerns especially with regard to the research undertakings of a specific type of organizations, namely Big Tech platforms. The latter have moved decisively to set foothold in the health domain—even through the attempted provision of healthcare services (e.g. Wingfield et al., 2018)—and have emerged as key enablers of data-intensive health research (Sharon, 2016, 2018). In so doing, however, they have proved to raise distinctive regulatory, ethical and socio-political challenges—which we review in the next section—which makes it highly questionable whether they should be allowed to benefit from the full spectrum of data protection facilitations afforded to scientific research undertakings.



## Big Tech platforms in health research

At the present day, pointing to the growing dominance of Big Tech corporations within an increasingly digitized health sector has arguably become a sort of truism. From research funding (e.g. Chan Zuckerberg Initiative, 2020) to the provision of research infrastructures (e.g. Broad Institute, 2015), from direct involvement in big data research (e.g. Verily, 2020) to the provision of technologies for recruitment of research participants, Big Tech has become a force to be reckoned with in health research.

The drawbacks stemming from the apparently inexorable rise of these powerful corporate actors in health research become visible at different levels. For one thing, Big Tech has long since being associated with dubious privacy-infringing practices. Aside from sanctioned data breaches and public scandals that have surfaced over the years, such as those involving Alphabet-controlled DeepMind in the UK (Powles and Hodson, 2017) and IBM Watson in Italy (Barbacetto, 2017), these corporations are known to routinely engage in ethically suboptimal yet legal practices predicated on large-scale data harvesting and the capacity to combine health, lifestyle, and mundane data types (e.g. search data) in myriad alarming ways (Fussell, 2019). These, in turn, can have a number of detrimental consequences for individual people as well as societies, ranging from breach of individual and group privacy (Mittelstadt, 2017) to social sorting (Hogle, 2016), from profiling (Hildebrandt and Gutwirth, 2008) to algorithmic discrimination (O’Neil, 2017), from increased inequality and social exclusion (Eubanks, 2018) to commercial “dataveillance” (Pasquale, 2015; Zuboff, 2019).

These “quasi legal” practices, and the ethical and social disruptions they entail, can be often traced to the existence of legal gray areas in data protection regimes, which makes it difficult to ascertain and/or clearly demonstrate breaches (McMahon et al., 2019). More in general, regulatory regimes, including the GDPR in the EU, reveal shortcomings in adequately governing digital health technologies and, notably, the data processing activities of digital health platforms, due to the challenges faced by data protection standards dating back to a pre-digital world (Marelli et al., 2020; Price and Cohen, 2019; Wachter, 2019).

Yet, while privacy and regulatory concerns have come to the fore in public debates, equally pressing are the broader societal issues raised by the very core architecture and business models of these organizations, which are able to achieve dominant positions in the markets in which they operate (Srnicek, 2017; Van Dijck et al., 2018). In particular, as they become integral part of leading research undertakings across the globe, Big Tech corporations are able to convert the advantage acquired in

the digital domain—through their technical expertise in terms of data collection, data analytics, and infrastructure development—into epistemic and governance advantages in the domain of health and medicine (Sharon, 2020; Shaw, 2020). As they increasingly engage established actors (academia, pharma) in research activities, these corporations not only manage to bring precious epistemic expertise in-house, thus emerging as prominent actors in research; also, they render their own digital expertise an indispensable “entry ticket” to new data-intensive fields of health research (Sharon, 2020). Moreover, by investing in the construction of repositories of public health, fitness, genomics, and health record data, they can easily become dominant data brokers controlling and establishing the rule of access to large-scale databases (Shaw, 2020). In turn, this gatekeeping function in health research translates into a governance advantage, as it grants these actors the power to reshape this domain according to their values and interests, while being able to define the future directions that the sector will take (Sharon, 2020). The agenda-setting prerogatives of Big Tech platforms are further reinforced by their capacity to target desired research areas for funding (Prainsack, 2020), establish close ties with key opinion leaders in the field (Shaw, 2020), and provide support to major science policy initiatives (e.g. Robbins, 2020)—thus exerting significant influence on standard- and policy-making processes. In sum, as Big Tech platforms strive in health research, they are increasingly poised to become obligatory passage points in the digitized biomedical landscape of the coming decades.

For all of the above, it appears questionable whether these actors should be made to benefit from the full spectrum of facilitations that the current data governance regime in Europe, centered around the GDPR, but involving also a relevant role by individual Member States (Donnelly and McDonagh, 2019; European Commission, 2021), renders available for scientific research. When derogations to stringent data protection requirements for data retention, secondary research, or information to data subjects are availed by Big Tech corporations—for instance, as it is routinely the case, within collaborative research networks—they could lead to a heightened risk of infringements to the rights and interests of individual people or vulnerable social groups. At the same time, these dynamics can have a lasting impact on the very social contract that underpins science–society relations, as well as reconfiguring the values in which European health systems are rooted.

## Toward a use-based and functionalist framework for scientific research

In this section, we review proposals and sketch out a possible path to unpack the GDPR's broadly construed notion of "scientific research" and functionally demarcate types of scientific research activities and data uses that can legitimately aspire to the regulatory leeway afforded by the GDPR and national legislation, and those instead to subject to higher data protection requirements. We do not intend to provide a full-fledged "governance fix" to address the issues identified above, yet we aim to outline points to consider that can be then further translated into more defined policy guidance for data-intensive health research undertakings (involving Big Tech platforms).

Moving from concerns similar to the ones articulated here, proposals have been made to steer the data governance approach toward *use-based data governance frameworks* (see especially Mantelero, 2014; Mayer-Schönberger and Padova, 2016; Prainsack, 2019), which are geared to distinguish desirable from undesirable data uses, and, on that basis, tailor facilitation mechanisms to the former, while subjecting the latter to higher data protection standards. Notably, weaving a use-based approach—one akin to those that have long since been in place in other sectors (e.g. pharmaceutical regulation)<sup>7</sup>—within the current GDPR-based legal regime could enable to more effectively *re-purpose* the data governance machinery toward the following aims.<sup>8</sup>

First, a use-based approach could enable to shift the focus of the data protection machinery from the informational self-determination of data subjects<sup>9</sup> toward the *ex ante* identification of desirable and undesirable personal data *uses* (Mayer-Schönberger and Padova, 2016; see also Cate and Mayer-Schönberger, 2013; Mantelero, 2014). In so doing, a use-based framework is geared to move upstream the assessment of potential criticalities deriving from personal data processing, taking such responsibility away from the by now largely fictitious construct of the "informed" and "consenting" data subject and bestowing it upon data governance bodies. In addition, it could also account for the *broader societal implications* of big data processing that the informational self-determination model—centered on the individual citizen, mostly *qua* consumer—largely eschew. As such, it would enable to increase *collective*, other than individual, *control* (Prainsack, 2017), thus providing the means to account for, and mitigate, the most harmful and socially disruptive practices carried out by Big Tech platforms, all the while still managing to incentivize research deemed beneficial for citizens and society.

The adoption of a use-based framework, centered on the concrete assessment of data uses, would also be conducive to the elaboration of a more fine-grained definition of scientific research and, relatedly, to arrive to a functionalist demarcation between research activities that should be made amenable to data protection incentives and those that should not.

Moving largely in this vein, Floridi and colleagues (2019) have proposed the notion of "*bona fide* research", whereby "research qualifies as *bona fide* whenever its ultimate goal is to discover new knowledge intended for the general interest in health and to be made publicly accessible (e.g., published in scientific journals or disseminated through digital media) without undue delay" (Floridi et al., 2019: 361). By referring to the "general interest", the authors aim to maintain a "lay and broad approach", avoiding a too narrow and substantive notion of what the public good should amount to, while, at the same time, preventing uses of health data that can be seen as *prima facie* undesirable, such as in military research, or serving specific interests, such as market research and intelligence gathering potentially exploited for targeted advertising (Floridi et al., 2019). In addition, *bona fide* research should be carried out by "*bona fide* research organizations", defined as "any organization appointed or accredited or funded to undertake bona fide research, and/or which has made public its commitment to adhere to recognized research governance principles" (Floridi et al., 2019: 365). Importantly, to count as a *bona fide* research organization, "it is also not a requirement that *bona fide* research is the primary business of that organization" (Floridi et al., 2019: 365), as the focus in the definition is on the specific *function* of the organization rather than on its *nature*.

The proposal by Floridi and colleagues is valuable in a number of respects. First, it explicitly recognizes the need for inscribing further specifications in the category of "scientific research". In addition, by maintaining a *functionalist* approach, and focusing on what research organizations *do* instead of what they *are*, it avoids simplistic dichotomies and binary framings (such as public vs private) that, as we observed above, do not largely stand the test of contemporary research practices. Moreover, it moves away from ill-suited considerations on sectoral epistemic standards, to focus instead on the normative orientation and governance context of research, which can be arguably considered more relevant elements when distinguishing between research activities to be incentivized or not (Boers et al., 2015), in line with empirical studies revealing that public propensity to engage in research is only to a limited extent affected by its specific epistemic details (Grady et al., 2015).<sup>10</sup>

Yet, such a definition faces challenges in applying to data-intensive research activities involving Big Tech



platforms. First, the framing of its goal as the “discovery of new knowledge” seems to corroborate, albeit even inadvertently, enduring distinctions between discovery and application, basic and translational that have increasingly lost sharpness in contemporary research practice, most certainly in data-intensive biomedicine. Additionally, the much welcome functionalist turn in “certifying” the *bona fide* research undertakings of an organization, regardless of its public or corporate nature, risks falling short when it comes to Big Tech platforms for which the conflation of roles is part and parcel of their very business model. Engaging with such organizations as “*bona fide* research” partners would then require additional provisions of clear demarcations within their activities so as to be compatible with the requirements of “general interest”. Relatedly, such demarcations appears of difficult actionability in relation to algorithmic training, in light of the potential re-purposing of health data, but also the “re-purposing” of the very technology (e.g. hardware or software alike) developed through such data for other types of research or clinical purposes.

Fourth, such a broad-construed framing of “general interest” fails to explicitly capture the multiple “normative repertoires” (from accelerating scientific innovation, to improving public health, to enhancing wealth creation, etc.) that drive, and provide social legitimacy and ethical justification to, contemporary data-intensive health research undertakings (Sharon, 2018). In turn, disregarding these different and in fact competing articulations of the public good can lead to overlook the fact that different actors involved in health research, as e.g. research participants or data custodians, may each hold different plausible conceptions of what constitutes “socially acceptable” types of research and data uses.<sup>11</sup> The problem is thus that if *any* notion of the common good may suffice to make research activities socially desirable, what such notion amounts to in practice may vary greatly and end up underlying plausibly divergent repertoires of justification (Sharon, 2018).

Accordingly, the implementation of a narrower yet sufficiently actionable notion of research undertakings and data uses eligible for data protection facilitations seems to require the consideration of the following additional elements.

In the first place, use-based frameworks should incorporate, and ensure the application of, strong mechanisms for preventing violations of “contextual integrity”<sup>12</sup> (Nissenbaum, 2004) throughout the full life cycle of personal data. This entails not only avoiding that health data is repurposed *outside* the health domain, such as for targeted advertising, but also, that any secondary processing is attuned to, and geared toward, the *distinctive* notion of the common

good underpinning the research purpose for which data have been originally collected. This can be traced, for instance, in the mission statement of the organization acting as data custodian, and further specified and fine-tuned in the data governance model drafted to regulate primary and secondary data processing activities. It should be noted that we are not advocating here for a narrow, technical notion of “purpose limitation” as enshrined in the GDPR, which we have claimed to be of difficult actionability in the present research landscape. Rather, we are proposing that secondary research, whatever its specific purpose may be, should be compatible—in a broader normative, rather than narrow technical-legal sense—with the underlying notion of the public good of the primary data processing.

In practice, implementing such mechanisms would require data custodians and data governance bodies (see below) to elaborate, on a context-sensitive basis, well-defined criteria for permitting data re-purposing while preserving contextual integrity, to be enforced through binding data transfer agreements (DTAs). In particular, DTAs should contain clauses permitting reuse of data within the facilitated data protection regime for types of research activities that pursue (context-related) forms of common good and do not involve increased risks for either research participants, larger social groups, or society as a whole.

An additional requirement for effective implementation of a finer-grained definition of scientific research undertakings amenable to data protection facilitations within a use-based model is the establishment of a new layer of regulatory bodies in the guise of *data governance boards* (DGBs) ubicated within each research institution and research networks. As a recent Europe-wide survey has shown (European Commission, 2021), a number of institutions and research networks still lack data governance bodies of this kind. These could play an important gatekeeping function in overseeing data processing and transfers, account for the social desirability of research, while operating the two-layered system of incentives and disincentives for different types of research. Under criteria of independence akin to those of ethics committees, DGBs would be in charge of overseeing and approving all data processing operations and data flows within and outside the organization through a multi-tier use-based system that streamlines or constraints accordingly. They would thus be responsible for exercising data governance throughout the full life cycle of personal data (Jacobs and Poma, 2019), while aiming to valorize personal data processing for research in compliance with ethics and normative requirements. Importantly, DGBs would also be tasked with performing an additional twofold function.

First, to enforce and carry out *effective auditing procedures* to monitor compliance, especially when data is transferred to opaque entities such as Big Tech platforms. While facing non-trivial organizational barriers, this requires a shift in common perception and *modus operandi* from the part of established actors in health research. Nonetheless, we argue, this shift will be key for achieving effective and socially robust governance of contemporary data-intensive research practices—notably those involving Big Tech platforms.

Second, such bodies would be tasked with implementing effective mechanisms of *benefit-sharing*, which should be seen as an essential complement to mechanisms for the preservation of contextual integrity. Mechanisms of benefit-sharing are paramount to ensure that processing of personal data is put to the service of the common good (however variously conceived this may be). Rather than merely guaranteeing returns to research participants (Hayden, 2007), such as those foreseen in the post-study obligations of clinical studies, these mechanisms should be geared to avoid depletion of public assets and technological know-hows, especially when publicly-collected data is transferred to powerful entities such as Big Tech platforms. In many cases, even if successfully implemented through binding DTAs, mechanisms for preserving contextual integrity would still be blunt in preventing dominant actors from accruing an arguably illegitimate competitive advantage by exploiting publicly-collected resources. Therefore, mechanisms of benefit-sharing (such as free-of-cost provision of proprietary technology developed through publicly collected data processing) would represent a means to fairly balance the distinct values and interests at stake in health research. Notably, as we observed above, since Big Tech platforms are able to accrue advantages by gathering data across distinct research projects, and by repurposing not just personal data but also proprietary technology, DGBs—especially those within large (national) research networks—would be in a position to exert a broad-spanning overseeing function, thus partially addressing the limitations of data governance mechanisms centered on individual research projects. In that regard, it is important to recognize that clinical centers and research institutions, in many cases undergoing forms of governmental accreditation (and thus being *lato sensu* part of “public” health systems), still represent a fundamental and often the main source for the collection of personal health data, and thus have negotiating powers that should not be relinquished when it comes to agreeing on data governance agreements with corporate actors.

Whilst unlikely to exhaust all issues triggered by the involvement of Big Tech in health research, these policy measures could help tame some of their most disruptive

impacts in personal data processing while chaperoning an orderly entry into the health research ecosystem.

## Conclusions

Since its entry into force in 2018, the GDPR has had an increasingly structuring effect on health research, in Europe and beyond. In this contribution, we analyzed the GDPR’s special derogatory regime for research in relation to Big Tech platforms inasmuch as, in light of the governance and social challenges they raise, these entities can be seen as representing the ultimate litmus test for a regulation designed with the explicit intent of achieving “the right balance” between a high level of protection for the fundamental right to data protection and the enhancement of digital innovation in the continent (Albrecht, 2016).

As it is by now evident, Big Tech platforms pose specific governance challenges and raise distinctive socio-political concerns in the health domain, revolving around monopoly/market-dominance, enclosure, and privatization dynamics. For sure, these dynamics do not arise out of, and do not owe exclusively to, the data protection derogations foreseen for scientific research. Nor amending and tightening the latter will automatically tame the disruptive impact of such platforms. Still, the special derogatory regime for research is poised to amplify—rather than contain—these concerns, and can thus risk providing a “regulatory back-door” that exacerbates some of the challenges faced by the GDPR in regulating digital health undertakings.

Moving from these premises, this contribution has built on a growing discussion in data protection and science policy debates to outline points to consider for the implementation of use-based data governance frameworks that tailor the scope and reach of the facilitations mechanisms for research to the ethical and societal desirability that different types of health research undertakings entail. Such a use-based approach is meant to move upstream the assessment of potential data protection criticalities, while also accounting for the broader societal implications of data-intensive health research undertakings. Specifically, it unpacks the abstract and broadly construed notion of “scientific research” advanced in the GDPR in order to enable the tracing of new taxonomies and demarcations within this fuzzy category. At the same time, we have stressed the importance of foreseeing mechanisms for preserving contextual integrity and providing benefit-sharing, so as to let the promissory research involving Big Tech platforms unleash its full potential while remaining tightly accountable to the multiple societal expectations associated with it.

In conclusion, all of the above may require a *repurposing* the data protection and data governance

machinery itself, departing from a narrow and increasingly unsatisfactory focus on privacy alone, so as to cater to the increasingly salient ethical and socio-political implications associated with big data processing in health research.

## Acknowledgements

The authors would like to thank the Editors and three anonymous reviewers for their insightful and constructive comments that helped strengthening the argument of the article. We would also like to acknowledge organizers and participants to the “*Solidarity Retreat*” (Vienna, 16–17 May 2019) and the workshop “*The Clinic and the Bank – Towards a Bioethics of the Information Revolution in Medicine*” (Van Leer Jerusalem Institute, 7–10 July 2019), as well as Gert Meyers (Tilburg University), for useful feedback on preliminary conceptualizations and earlier drafts of the article.




## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

This work was supported by the European Union’s Horizon 2020 research and innovation program, under the Marie Skłodowska-Curie grant agreement number 753531 (LM).

## ORCID iDs

Luca Marelli  <https://orcid.org/0000-0002-3960-0033>  
 Giuseppe Testa  <https://orcid.org/0000-0002-9104-0918>  
 Ine Van Hoyweghen  <https://orcid.org/0000-0002-9402-2918>

## Notes

1. Following Van Dijck et al. (2018: 4), platforms can be defined as “programmable digital architecture[s] designed to organize interactions between users”, geared toward “the systematic collection, algorithmic processing, circulation, and monetization of users data”. While these features intuitively connote (entirely virtual) eHealth and mHealth platforms, they also apply to data repositories, cloud computing platforms, and complex human tissue product platforms generating and processing large amounts of data, such as, for instance, stem cell and organoid platforms (Marelli and Testa, 2019).
2. “Such negotiations are part of the so-called ordinary legislative procedure of the EU, cf.: <https://europarl.europa.eu/ordinary-legislative-procedure/en/interinstitutional-negotiations.html>
3. The question as to the relation between Art. 6 and Art. 9 is open to distinct interpretations. Here, we read Article 9 (2)(j) in conjunction with the standards for lawful processing under Art. 6(1). For a more in-depth discussion on this point, see Article 29 Working Party (2014), Donnelly and McDonagh (2019: 112), and especially Bolognini et al. (2016).
4. However, the alleged difficulties posed by the GDPR in cross-border data transfers with third countries such as the US, notably within international research collaborations, have been the subject of increased critical remarks (Bovenberg et al., 2020; Peloquin et al., 2020; Rabesandratana, 2019; Scheibner et al., 2020).
5. For a more in-depth discussion of this point, cf. the recent guidelines on health data processing for the purpose of scientific research in the context of the COVID-19 outbreak (EDPB, 2020: 8–11).
6. Personal communications with Members of different Ethics Committees and Research Ethics Consultants in Italy.
7. Use-based regulations have long since been in place in other sectors (e.g. food, aviation, pharmaceutical) that, at once, entail negative externalities (e.g. drugs’ adverse effects) and are too complex for individual people to navigate without prior expert knowledge (Mayer-Schönberger and Padova, 2016). Limiting the space for autonomous risk-assessment by individual people (which is confined, for instance, to off-label drug uses), regulation in these domains is instead geared to preempt the occurrence of individual or societal harms.
8. Notably, the GDPR affords a twofold pathway for “regulatory adjustments”, either through national legislation (cf. e.g. Art. 9(2)(j) and 9(4) GDPR) or soft-rule instruments (e.g. codes of conduct, Art. 40 GDPR) that tailor the implementation of the Regulation to specific processing sectors, such as health research. Therefore, through either route, it is possible to devise data governance models that can be weaved within the current GDPR-based legal regime.
9. For a more elaborated discussion on the pitfalls of the informational self-determination approach in current digital environments, see e.g. Marelli et al., 2020.
10. At the same time, it is important to stress that research that can legitimately aspire to benefit from the facilitated data protection regime should be conducted on the basis of a minimum set of commonly agreed epistemic criteria. These may include a robust scientific rationale; reproducibility; compliance with recognized criteria for scientific integrity; on this, see COREON, 2019, Van Veen, 2018.
11. For instance, while some may consider research undertakings purposed toward advancing economic growth through digital innovation as a type of research that can be subsumed under the rubric of the “general interest” or the “common good”, some others may hold strong competing views on this point.
12. Violations of contextual integrity are defined by Nissenbaum as violations of context-relative and socially shared “informational norms”, notably, for our discussion, “norms of distribution”, which govern “the flow or distribution of information – movement, or transfer of information from one party to another or others” (Nissenbaum, 2004: 122).



## References

- Abbott A (2015) European medical research escapes stifling privacy laws. Proposed legislation had threatened the use of genomic and clinical data in medical studies. *Nature News* 16: 1.
- Albrecht J (2016) Conclusion of the EU data protection reform. Available at: <https://janalbrecht.eu/2016/04/2016-04-13-conclusion-of-the-eu-data-protection-reform/> (accessed 14 October 2020).
- Article 29 Working Party (2013) Opinion 03/2013 on purpose limitation. Adopted on 2 April 2013. 00569/13/EN WP 203.
- Article 29 Working Party (2014) Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC. Adopted on 9 April 2014. 844/14/EN WP 217.
- Article 29 Working Party (2017) Guidelines on Consent under Regulation 2016/679. Adopted on 28 November 2017. Last Revised and Adopted on 10 April 2018. WP259 rev.01.
- Barbacetto G (2017) A Ibm tutti i nostri dati sanitari. In cambio della nuova sede sull'area Expo. Available at: <http://giannibarbacetto.it/2017/02/15/a-ibm-tutti-i-nostri-dati-sanitari-in-cambio-della-nuova-sede-sullarea-expo/> (accessed 14 October 2020).
- Boers SN, Van Delden JJ and Bredenoord AL (2015) Broad consent is consent for governance. *The American Journal of Bioethics : AJOB* 15(9): 53–55.
- Bolognini L, Pellino E and Bistolfi C (2016) *Il regolamento privacy europeo: commentario alla nuova disciplina sulla protezione dei dati personali*. Milan, Italy: Giuffrè editore.
- Bovenberg J, Peloquin D, Bierer B, et al. (2020) How to fix the GDPR's frustration of global biomedical research. *Science (New York, N.Y.)* 370(6512): 40–42.
- Broad Institute (2015) Broad Institute, Google Genomics combine bioinformatics and computing expertise to expand access to research tools. Press release 23 June 2015. Available at: <https://broadinstitute.org/news/broad-institute-google-genomics-combine-bioinformatics-and-computing-expertise-expand-access> (accessed 14 October 2020).
- Burrell J (2016) How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society* 3(1).
- Buttarelli G (2018) Speech at the Fifth World Congress for Freedom of Scientific research. 12 April 2018. Available at: [https://edps.europa.eu/sites/edp/files/publication/18-04-12\\_fifth\\_world\\_congress\\_freedom\\_scientific\\_research\\_en.pdf](https://edps.europa.eu/sites/edp/files/publication/18-04-12_fifth_world_congress_freedom_scientific_research_en.pdf) (accessed 14 October 2020).
- Cambrosio A, Keating P, Schlich T, et al. (2006) Regulatory objectivity and the generation and management of evidence in medicine. *Social Science & Medicine* (1982) 63(1): 189–199.
- Carrier M and Nordmann A (2011) Science in the context of application: Methodological change, conceptual transformation, cultural reorientation. In: Carrier M and Nordmann A (eds) *Science in the Context of Application*. Dordrecht, the Netherlands: Springer, pp.1–7.
- Casali P (2014) Risks of the new EU data protection regulation: An ESMO position paper endorsed by the European oncology community. *Annals of Oncology: Official Journal of the European Society for Medical Oncology* 25: 1458–1461.
- Cate FH and Mayer-Schönberger V (2013) Notice and consent in a world of big data. *International Data Privacy Law* 3 (2): 67–73.
- Chan Zuckerberg Initiative (2020) Grants. Available at: <https://chanzuckerberg.com/justice-opportunity/> (accessed 14 October 2020).
- Chivot E (2019) One year on, GDPR needs a reality check. *Financial Times*, 30 June. Available at: <https://ft.com/content/26ee4f7c-982d-11e9-98b9-e38c177b152f> (accessed 14 October 2020).
- Cohen J (2019) Google, Ascension data partnership sparks federal probe. *Modern Healthcare*. Available at: <https://modernhealthcare.com/information-technology/google-ascension-data-partnership-sparks-federal-probe> (accessed 14 October 2020).
- Conrado E (2019) Technology that improves patients' lives, caregivers' experience. Available at: <https://ascension.org/News/News-Articles/2019/11/12/21/45/Technology-that-improve-patients-lives-caregivers-experience> (accessed 14 October 2020).
- Copeland R and Needleman SE (2019) Google's 'project nightingale' triggers federal inquiry. *Wall Street Journal*, 12 November. Available at: <https://wsj.com/articles/behind-googles-project-nightingale-a-health-data-gold-mine-of-50-million-patients-11573571867> (accessed 14 October 2020).
- Copeland R, Mattioli D and Evans M (2020) Paging Dr. Google: How the tech giant is laying claim to health data. *Wall Street Journal*, 11 January. Available at: <https://wsj.com/articles/paging-dr-google-how-the-tech-giant-is-laying-claim-to-health-data-11578719700> (accessed 14 October 2020).
- COREON (2019) Statement on scientific research. Available at: [https://federa.org/sites/default/files/images/statement\\_scientific\\_research\\_28jul2019.pdf](https://federa.org/sites/default/files/images/statement_scientific_research_28jul2019.pdf) (accessed 14 October 2020).
- Decreto Legislativo 10 agosto 2018, n. 101, Disposizioni per l'adeguamento della normativa nazionale alle disposizioni del regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati). (18G00129) (GU Serie Generale n.205 del 04-09-2018).
- Donnelly M and McDonagh M (2019) Health research, consent and the GDPR exemption. *European Journal of Health Law* 26(2): 97–119.
- Dove ES, Thompson B and Knoppers BM (2016) A step forward for data protection and biomedical research. *The Lancet* 387(10026): 1374–1375.
- Eubanks V (2018) *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor*. New York, NY: St Martin's Press.

- European Commission (2016) Goals of research and innovation policy. Available at: [https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy\\_en](https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy_en) (accessed 14 October 2020).
- European Commission (2021) Assessment of the EU Member States' rules on health data in the light of GDPR, Luxembourg: Publications Office of the European Union, 2021. Available at: [https://ec.europa.eu/health/sites/health/files/ehealth/docs/ms\\_rules\\_health-data\\_en.pdf](https://ec.europa.eu/health/sites/health/files/ehealth/docs/ms_rules_health-data_en.pdf) (accessed 1 March 2021).
- European Data Protection Board (EDPB) (2020) Guidelines 03/2020 on the processing of data concerning health for the purpose of scientific research in the context of the COVID-19 outbreak. Adopted on 21 April 2020. Available at: [https://edpb.europa.eu/sites/edpb/files/files/file1/edpb\\_guidelines\\_202003\\_healthdatascientificresearch\\_covid19\\_en.pdf](https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202003_healthdatascientificresearch_covid19_en.pdf) (accessed 14 October 2020).
- European Data Protection Board (EDPB) (2021) EDPB Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research. Adopted on 2 February 2021. Available at: [https://edpb.europa.eu/sites/edpb/files/files/file1/edpb\\_replyec\\_questionnaire\\_research\\_final.pdf](https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_replyec_questionnaire_research_final.pdf) (accessed 1 March 2021).
- European Data Protection Supervisor (EDPS) (2020) A Preliminary Opinion on data protection and scientific research. Available at: [https://edps.europa.eu/sites/edp/files/publication/20-01-06\\_opinion\\_research\\_en.pdf](https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf) (accessed 14 October 2020).
- Floridi L, Luetge C, Pagallo U, et al. (2019) Key ethical challenges in the European medical information framework. *Minds and Machines* 29(3): 355–371.
- Fussell S (2019) Google's totally creepy, totally legal health-data harvesting. *The Atlantic*, 14 November. Available at: <https://theatlantic.com/technology/archive/2019/11/google-project-nightingale-all-your-health-data/601999/> (accessed 14 October 2020).
- Grady C, Eckstein L, Berkman B, et al. (2015) Broad consent for research with biological samples: Workshop conclusions. *The American Journal of Bioethics: AJOB* 15(9): 34–42.
- Hallinan D (2020) Broad consent under the GDPR: An optimistic perspective on a bright future. *Life Sciences, Society and Policy* 16(1): 1–18.
- Hayden C (2007) Taking as giving: Bioscience, exchange, and the politics of benefit-sharing. *Social Studies of Science* 37(5): 729–758.
- Hildebrandt M and Gutwirth S (2008) *Profiling the European Citizen*. Dordrecht, the Netherlands: Springer.
- Hoeyer K (2016) Denmark at a crossroad? Intensified data sourcing in a research radical country. *The Ethics of Biomedical Big Data* 29: 73–93.
- Hogle LF (2016) Data-intensive resourcing in healthcare. *BioSocieties* 11 (3): 372–393.
- Jacobs B and Popma J (2019) Medical research, big data and the need for privacy by design. *Big Data & Society* 6(1).
- Keating P and Cambrosio A (2003) *Biomedical Platforms: Realigning the Normal and the Pathological in Late-Twentieth-Century Medicine*. Cambridge, MA: Mit Press.
- LERU (2016) The new EU General Data Protection Regulation: Why it worries universities and researchers. Policy Brief 14 April 2016. Available at: <https://leru.org/news/the-new-eu-general-data-protection-regulation-why-it-worries-universities-researchers> (accessed 14 October 2020).
- Mager A (2017) Search engine imaginary: Visions and values in the co-production of search technology and Europe. *Social Studies of Science* 47(2): 240–262.
- Mantelero A (2014) The future of consumer data protection in the E.U. Re-Thinking the “notice and consent” paradigm in the new era of predictive analytics. *Computer Law & Security Review* 30(6): 643–660.
- Marelli L and Testa G (2017) ‘Having a structuring effect on Europe’: The innovative medicines initiative and the construction of the European bioeconomy. In: Goven J and Pavone V (eds) *Bioeconomies: Life, Technology and Capital in the XXIst Century*. London, UK: Palgrave.
- Marelli L and Testa G (2018) Scrutinizing the EU general data protection regulation. *Science (New York, N.Y.)* 360(6388): 496–498.
- Marelli L and Testa G (2019) iPSC- and organoid-based biomedicine at the intersection of epigenetics and regeneration: Charting the normative contours of emerging biomedical platforms. In: Palacios D (ed.) *Epigenetics and Regeneration*. Cambridge, MA: Academic Press.
- Marelli L, Lievevrouw E and Van Hoyweghen I (2020) Fit for purpose? The GDPR and the governance of European digital health. *Policy Studies* 41(5): 447–467.
- Mayer-Schönberger V and Padova Y (2016) Regime change? Enabling big data through Europe's new data protection regulation. *The Columbia Science & Technology Law Review* 17(1): 315–335.
- McDermott M, Wang S, Marinsek N, et al. (2019) Reproducibility in machine learning for health. *arXiv preprint arXiv:1907.01463*.
- McMahon A, Buyx A and Prainsack B (2019) Big data governance needs more collective responsibility: The role of harm mitigation in the governance of data use in medicine and Beyond. *Medical Law Review* 28(1): 155–182.
- Mittelstadt B (2017) From individual to group privacy in big data analytics. *Philosophy & Technology* 30(4): 475–494.
- Nissenbaum H (2004) Privacy as contextual integrity. *Washington Law Review* 79(1): 119–157.
- Nowothny H, Scott P and Gibbons M (2001) *Re-Thinking Science: Knowledge and the Public in the Age of Uncertainty*. Hoboken, NJ: Wiley.
- O'Neil C (2017) *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy*. New York, NY: Penguin Books.
- Pasquale F (2015) *The Black Box Society*. Cambridge, MA: Harvard University Press.
- Peloquin D, Di Maio M, Bierer B, et al. (2020) Disruptive and avoidable: GDPR challenges to secondary research uses of data. *European Journal of Human Genetics* 28: 697–705.

- Ploem MC, Essink-Bot ML and Stronks K (2013) Proposed EU data protection regulation is a threat to medical research. *BMJ* 345.
- Pormeister K (2017) Genetic data and the research exemption: Is the GDPR going too far? *International Data Privacy Law* 7(2): 137–146.
- Powles J and Hodson H (2017) Google DeepMind and healthcare in an age of algorithms. *Health and Technology* 7(4): 351–367.
- Prainsack B (2017) Research for personalised medicine: Time for solidarity. *Medicine and Law* 36(1): 87–98.
- Prainsack B (2019) Logged out: Ownership, exclusion and public value in the digital data and information commons. *Big Data & Society* 6(1).
- Prainsack B (2020) The political economy of digital data: Introduction to the special issue. *Policy Studies* 41(5): 439–446.
- Price WN and Cohen IG (2019) Privacy in the age of medical big data. *Nature Medicine* 25(1): 37–43.
- Rabesandratana T (2019) Researchers sound alarm on European data law. *Science (New York, N.Y.)* 366(6468): 936.
- Regulation (EU) 2016/679 (General Data Protection Regulation) 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, which repeals Directive 95/46/EC. Available at: [www.eugdpr.org/](http://www.eugdpr.org/) (accessed 7 May 2021).
- Robbins R (2020) The White House is pinning its hopes on health tech to save the day. Can it deliver? *STAT*, 18 March. Available at: <https://statnews.com/2020/03/18/coronavirus-white-house-pinning-hopes-on-health-tech/> (accessed 14 October 2020).
- Scheibner J, Ienca M, Kechagia S, et al. (2020) Data protection and ethics requirements for multisite research with health data: A comparative examination of legislative governance frameworks and the role of data protection technologies. *Journal of Law and the Biosciences* 7(1).
- Science Europe (2013) Position Statement on the Proposed European General Data Protection Regulation. May 2013.
- Shabani M and Marelli L (2019) Re-identifiability of genomic data and the GDPR: Assessing the re-identifiability of genomic data in light of the EU general data protection regulation. *EMBO Reports* 20(6): e48316.
- Shabani M, Chassang G and Marelli L (2021) The impact of the GDPR on the Governance of Biobank research. In: Slokenberga S, Tzortzatou O and Reichel J (eds) *GDPR and Biobanking: Individual Rights, Public Interest and Research Regulation across Europe*. Berlin, Germany: Springer, pp.45–60.
- Sharon T (2016) The googlization of health research: From disruptive innovation to disruptive ethics. *Personalized Medicine* 13(6): 563–574.
- Sharon T (2018) When digital health meets digital capitalism, how many common goods are at stake? *Big Data & Society* 5(2).
- Sharon T (2020) Blind-sided by privacy? Digital contact tracing, the Apple/Google API and Big Tech's newfound role as global health policy makers. *Ethics and Information Technology*. Epub ahead of print, <https://doi.org/10.1007/s10676-020-09547-x>
- Shaukat E (2019) Our partnership with Ascension. In Google Cloud Blog, 11 December. Available at: <https://cloud.google.com/blog/topics/inside-google-cloud/our-partnership-with-ascension> (accessed 14 October 2020).
- Shaw JA (2020) Policy and governance of artificial intelligence for health: A global ethics perspective. In: *Pontifical academy for life conference the "Good" Algorithm? Artificial intelligence ethics, law, health*. Vatican City, 26–28 February 2020.
- Srnicek N (2017) *Platform Capitalism*. Hoboken, NJ: John Wiley & Sons.
- Starkbaum J and Felt U (2019) Negotiating the reuse of health-data: Research, big data, and the European general data protection regulation. *Big Data & Society* 6(2).
- Staunton C, Slokenberga S and Mascalzoni D (2019) The GDPR and the research exemption: Considerations on the necessary safeguards for research biobanks. *European Journal of Human Genetics: EJHG* 27(8): 1159–1167.
- Taylor L and Purtova N (2019) What is responsible and sustainable data science? *Big Data & Society* 6(2).
- Thornhill J (2018) There is a 'third way' for Europe to navigate the digital world. *Financial Times*, 19 November. Available at: <https://ft.com/content/9da4156c-ebd4-11e8-89c8-d36339d835c0> (accessed 14 October 2020).
- Van Dijk J, Poell T and De Waal M (2018) *The Platform Society: Public Values in a Connective World*. Oxford, UK: Oxford University Press.
- van Veen EB (2018) Observational health research in Europe: Understanding the general data protection regulation and underlying debate. *European Journal of Cancer (Oxford, England: 1990)* 104: 70–80.
- Verily (2020) Collaborating with Vanderbilt and the Broad Institute to deliver the infrastructure and research tools for All of Us. Available at: <https://verily.com/solutions/all-of-us/> (accessed 14 October 2020).
- Verschraegen G (2018) Regulating scientific research: A constitutional moment? *Journal of Law and Society* 45: S163–S184.
- Wachter S (2019) Data protection in the age of big data. *Nature Electronics* 2(1): 6–7.
- Wingfield N, Thomas K and Abelson R (2018) Amazon, Berkshire Hathaway and JPMorgan team up to try to disrupt health care. *The New York Times*, 30 January. Available at: <https://nytimes.com/2018/01/30/technology/amazon-berkshire-hathaway-jpmorgan-health-care.html> (accessed 14 October 2020).
- Zarsky T (2017) Incompatible: The GDPR in the Age of Big Data. *Seton Hall Law Review* 47(4): 995–1020.
- Zuboff S (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London, UK: Profile Books.