

CHAPTER 6

Emerging (Inter-)Dependencies and their Implications

In big data-driven health research, entanglements between academic studies and market-dominating tech/internet corporations have emerged. This is in part related to the tendency that access to online data is increasingly controlled by these companies. Research projects drawing on, for example, social media data depend on collection and access conditions defined by internet and tech corporations. This is also linked, however, to tech corporations' philanthro-capitalist engagement in funding and encouraging research at the intersection of public health and tech-driven innovation. Tech-related topics, development and data science approaches in health research are supported through corporate data, analytics and grant schemes.

How data are retrieved by internet/tech corporations reflects certain norms and values. Big data-driven health research that uses data collected under corporate conditions, runs the risk of echoing and normalising these values and norms as they become decisive conditions for projects' data retrieval. In consequence, this research also reinforces the moral credibility of corporate approaches to users' data by showcasing big data's contribution to societal well-being and public health.

These tendencies have crucial implications for research ethics and integrity. It is particularly notable that studies involving big data tend to diminish possibilities through which affected actors could voice their (dis-)approval. Relevant stakeholders, in particular data subjects, are barely involved in negotiations of norms relevant to data retrieval or use. Informed consent is abandoned, mostly without questioning the appropriateness to do so for specific studies.

From a discourse ethics perspective, the validity of moral norms in big data-driven health research is assessed by asking how they were created in formative discourse (see e.g. Habermas 2001 [1993], 1990). Habermas proposes that the validity of norms depends on whether their assertion safeguards the autonomy

How to cite this book chapter:

Richterich, A. 2018. *The Big Data Agenda: Data Ethics and Critical Data Studies*.

Pp. 91–103. London: University of Westminster Press.

DOI: <https://doi.org/10.16997/book14.f>. License: CC-BY-NC-ND 4.0

of all affected individuals. As a ‘counterfactual idealization’ (Rehg 2015, 30), his theory is meant to guide and assess (moral) reasoning. ‘Justice’ is seen as a key dimension of validity for moral discourses; valid norms are those ensuring justice. Habermas’ theory has been frequently criticised as utopian. But even though its main normative principles may be ultimately out of reach, they provide reference points towards which (moral) reasoning may orient itself.

Addressing the validity of those social norms guiding big data-driven health research is highly relevant, as ethico-methodological changes in this field compromise many long-established research principles, such as informed consent. As described in Chapter 2, my analysis addresses two main issues concerning big data-driven health research, derived from critical data studies, pragmatist ethics and Habermasian theory: what are the broader discursive conditions, including key stakeholders and factors shaping their views? Which ethical arguments and validity claims have been brought forward? In this chapter, I reflect on the implications of observations and arguments presented in response to these questions in Chapters 4 and 5: stakeholders, discursive conditions and validity claims.

Stakeholders, Discursive Conditions, Validity Claims

Stakeholders

With regards to affected actors, I maintain that there is currently an imbalance and lack of formative discourse defining the ethics and social norms of big data-driven health research. Emerging data practices and ethics are criticised by academics and (occasionally) data activist groups, such as the Electronic Frontier Foundation. But often these debates are carried out in response to big data-driven approaches, rather than being foregrounded by involved researchers themselves. Moreover, there is little formative dialogue between researchers exploring novel approaches and those challenging ethical assumptions made with this research. There is also little discursive involvement when it comes to affected, civic individuals whose data are (or could be) used (Lupton 2016; Metcalf and Crawford 2016). This issue stresses the relevance of enhanced efforts in communicating relevant scientific developments and ethical dimensions of big data-driven research in public health domains.

Such efforts are crucial for fostering individuals’ possibilities to voice concern or approval. There is an urgent need to facilitate civic insights and possibilities for formative moral discourse regarding emerging, big data-driven research approaches. This observation also corresponds with what Kennedy and Moss (2015) conceptualise as a much-needed transition towards approaching data subjects as ‘knowing’ rather than merely ‘known publics’. The authors criticise current data practices for addressing publics mainly as passive data subjects, as they are primarily aimed at making sense of datafied individuals

(see also Zwitter 2014). Instead data should be used to '[...] help members of the public to understand public issues and each other better, such that more informed and knowing publics may take shape' (Kennedy and Moss 2015, 8). In the case of big data-driven health research such an understanding can only (potentially) occur if research methods are made available for debate in accessible and apprehensible ways. Such research might then also call attention to how personal and sensitive users' digital data really are.

Following Habermas' principles of discourse and universalisation, the only possibility to justify or counter norms which are decisive for big data-driven health research – for example the negligence of informed consent – is to ensure individuals' engagement in practical discourse. Without enhanced investments in involving affected individuals discursively, emerging possibilities for big data access amplify alienation between researchers using, and individuals contributing, data. In many cases, this implies a lessened involvement of affected individuals in relevant discourses and a weakened validity of the moral norms at the heart of such academic research. This is particularly noticeable when looking at debates concerning the role of informed consent.

Scholars involved in and observing big data research have controversially discussed whether the negligence of informed consent is indeed morally reasonable or merely technologically induced in big data-driven research. Informed consent is dismissed by those engaged in big data-driven research as superfluous for studying data subjects, as a relic of obsolete data retrieval conditions and as a now avoidable source of bias. For those defending informed consent, however, informed consent is an indispensable tool for safeguarding the autonomy and dignity of affected individuals. Undoubtedly, informed consent does not perfectly match Habermas' idealised principles and idea of formative discourse. Yet it functions as a research element aimed at approximating conditions for collectively formed, valid and just norms which are ethically decisive for scholarly practices.

By relinquishing informed consent, scholars remove means for involving individuals in a discourse of normative approval or disapproval. In this sense, studies using big data and eschewing informed consent lack forms of discursive involvement fostered in earlier research approaches. In Habermasian terms, such studies move further away from conditions facilitating valid norms '[...] that meet (or could meet) with the approval of all affected in their capacity as participants in a practical discourse' (1990, 66). Current big data-driven research approaches tend to cut out informed consent as an established form of discursive engagement of affected individuals. They also commonly fail to implement alternative possibilities for discursive negotiations of this moral norm.

One of the still rare cases in which such an attempt has been made is the study by Young et al. (n.d.). As described in Chapter 5, their project aims at creating a platform for monitoring tweets which may indicate health related high-risk behaviour in a population. At the same time though, they conduct

interviews with individuals working with HIV organizations, as well as participants affected by HIV, on ethical issues regarding the taken approach. As indicated above, whether such approaches are indeed an acceptable alternative to informed consent has been questioned. Nevertheless, such strategies indicate how alternative means for shaping the discursive conditions for public opinion formation and the involvement of affected individuals can be explored.

Discursive conditions

With regards to discursive conditions, I argue that by engaging in big data-driven health research without foregrounding potential risks and ethical issues, scholars facilitate discouragement of discursive, civic involvement. By failing to stress their awareness of potential controversies, they moreover risk scandalisation and increased public mistrust towards emerging, data-driven research approaches. Researchers present the use of big data from a societal position to which the highest moral standards are supposed to apply. They rely heavily on their perception as acting in the interest of the public (Van Dijck 2014). Public trust has been acknowledged as crucial to scientific research practices and moral values in democratic societies (Wynne 2006; Kelch 2002). When using certain kinds of big data in academic research, scholars assert the moral adequacy of norms relevant to their research. At the same time, they assert the appropriateness and value of (corporate) practices needed to acquire the used data.

Mobilising and drawing on the public trust which is widely placed in academic research,⁸⁹ they likewise suggest that public scrutiny of big data practices is not necessary. In doing so, however, they fail to facilitate a better public understanding of how personal and sensitive social media data may be. This both fosters the abovementioned negligence of stakeholders and in turn, weakens the validity of morals crucial to research. When ethical debates happen, they often have an effect on public trust in science. The importance of ethical foresight has therefore also been stressed with regards to avoiding a ‘whiplash effect’, i.e. (over-)regulations due to extremely negative perceptions of scientific and technological developments (Mittelstadt and Floridi 2016, 305ff.).⁹⁰

These risks are related to competitive funding systems for public health research in which not only governmental grant schemes, but internet and tech corporations have come to play a distinct role. I elaborated in Chapter 4 that internet/tech corporations engage in supporting and funding projects investigating how digital technologies and big data may be employed. They particularly target domains considered as beneficial and relevant to societal development, notably public health research. This also means that such companies play a role in shaping contemporary research agendas. These corporate funding opportunities incentivise studies exploring how technological developments more generally, and big data specifically, can be used in research. Furthermore, such

funding schemes, and especially research taking place within corporations, are not overseen under the same conditions as research funded through governmental grant schemes (concerning, for example, ethical review).

Significant interest in the intersection of technology and big data, science and public health does not only apply to corporate funding and support. Governmental, (inter-)national funding schemes reinforce investments in tech and big data-driven research. The need to acquire funding to conduct research is a common prerequisite for contemporary scholarship (Hicks 2012; Benner and Sandström 2000). The conditions, criteria and ramifications of governmental funding schemes have been widely criticised, though (Geuna 2001). Berezin (1998) even famously stated that '[a] random lottery among the competent applicants would do equally well and, perhaps, even better, because it at least avoids the bias of sticking to current fads and fashions so typical of the conventional APR of research proposals' (10). Moreover, the significance of lobbying and policy developments for research trends has been pointed out (Parsons 2004).

Yet while also being far from complying with the Habermasian ideal of discursive conditions taking into account all potentially affected individuals, in democratic societies, governmental funding schemes aim at reflecting democratic values and decision-making processes. In contrast, corporate funding instruments are part of the rise of philanthrocapitalism, and of what Horvath and Powell (2016) termed 'disruptive philanthropy' (89; see also Alba 2016). It is characteristic for internet and tech corporations engaged in philanthrocapitalist strategies to invest in projects promising to improve societal wellbeing through technological innovation.

Corporate interests and agendas, such as technology and its benefits, are merged with domains that are associated with widely accepted moral values, notably related to public health. In most of these cases, the charitably invested money will not be taxed in ways which would have led – at least partly – to its contributing to governmental programmes guided by democratic values (Alba 2016; Horvath and Powell 2016). When research funding is linked to corporate interests, efforts aimed at democratic decision-making processes concerning research grants and schemes are undermined. Not only interdependencies, but also dependencies and conflicts of interest emerge: corporations are providing data, analytics, interfaces and grants for studies that are relevant to their economic interests and public image.⁹¹ These dynamics raise the question to what extent tech corporate agendas are getting 'baked into' research projects.

Complex interdependencies emerge especially around those projects using data and tools from the tech corporations that fund them. Sharon (2016a) reminds us that '[...] insofar as the devices and services that generate, store, and in some cases analyze these data are owned by commercial entities that are outside traditional health care and research, we also should be attentive to new power asymmetries that may emerge in this space, and their implications for the shaping of future research agendas.' These constellations result

in dependencies and potential conflicts of interest which may be difficult for involved scientists to resolve. The issue also relates back to the abovementioned concerns that the merging of corporate data retrieval and academic research may be hazardous to the reputation of the latter.

Public-private partnerships, for example between university projects and tech corporations, affect the public perception of both. Corporations providing data or grants benefit from associating themselves with the relevance and contributions of scientific endeavours. At the same time, scientists may be increasingly associated with moral concerns pertinent to corporate practices. With regards to initiatives using big data, the UK Science and Technology Committee (2015) stresses that misuses and leaks of data have fostered public distrust towards governmental as well as corporate practices: referring to studies conducted by pressure groups such as Big Brother Watch Ltd., the report notes '[...] that 79% of adults in the UK were 'concerned' about their privacy online, and 46% believed that they were 'being harmed by the collection of their data by large companies' (Science and Technology Science and Technology Committee, House of Commons 2015).

These assessments partly contrast with a 2014 Eurobarometer survey on 'Public perception of science research and innovation' and the European Commission's report published on its results. In response to this report, Floridi (2014) summarises its main results and suggests possible interpretations:

'As a priority, data protection ranks as low as quality of housing: nice, but very far from essential. The authors [of the *Eurobarometer* report] quickly add that 'but this might change in the future if citizens are confronted with serious security problems'. They are right, but the point remains that, at the moment, all the fuss about privacy in the EU is a political rather than a social priority. [...] Perhaps we 'do not get it' when we should (a bit like the environmental issues) and need to be better informed. Or perhaps we are informed and still think that other issues are much more pressing.' (500)

This book emphasises the first-mentioned option, i.e. the lack of information and formative discourse. It stresses, moreover, that this notably applies to the disregarded ethical issues and wider societal implications of techno-social big data entanglements. For instance, as long as it remains underemphasised and unclear what ramifications a lack of data protection may have for public health and individual healthcare, important arguments needed for formative discourse are systematically excluded. From a Habermasian perspective, this is less an issue of 'not getting it', but rather a matter of shaping individuals' chances for appreciating an issue and voicing (dis-)approval.

In this context, interdependencies between science, public trust, societal hopes and expectations are of key importance. Van Dijck's work pointedly highlights the relevance of scientists as key pillars of social trust, its formation

and mobilisation: ‘a paradigm resting on the pillars of academic institutions often forms an arbiter of what counts as fact or opinion, as fact or projection’ (2014, 206). In this sense, scientists involved in big data-driven research lend credibility to the assumption that corporate tech data can make a much needed contribution to societal wellbeing, thus potentially justifying compromises regarding individual rights. They give credibility to the (questionable) assumption that corporate data collection approaches are morally indisputable and ethical debates hence unnecessary.

This likewise discourages public negotiations of big data practices, and impedes discursive conditions for which the ‘force of the better argument’ (Keulartz et al. 2004, 19) is decisive. A major reason for this is that criticism is implicitly framed as unnecessary and futile, as well as selfish and detrimental: unnecessary, since big data’s use in public health research asserts the moral appropriateness of corporate data retrieval; futile, since these approaches are authoritatively presented as already established technological and moral ‘state of the art’; and selfish and detrimental, considering normative claims for the societal benefits attributed to big data.

Therefore, discursive conditions for big data-driven health research and related norms urgently require amplified, research-driven efforts for facilitating public debate, and the involvement of affected individuals. Yet instead we are witnessing another instance and variation of the pacing problem (Marchant, Allenby and Herkert 2011). While technological innovation has been embraced in big data-driven public health research, scrutinising ethical issues has been largely eschewed, and learning from controversies hindered.

Validity claims

The involvement of data subjects is largely missing in ethical negotiations concerning big data-driven health research. However, normative arguments are brought forward by academics involved in or affected by such research. These discourses illustrate the validity claims through which big data-driven approaches are justified or opposed.

Scholars such as Rothstein and Shoben (2013) as well as Ioannidis (2013) vehemently oppose the argument that informed consent has become irrelevant in big data-driven research. In terms of validity claims, they reject this tendency by raising doubt as to the normative rightness as well as the accuracy of statements made by proponents of big data research. According to the authors, neglecting informed consent neither warrants the alleged methodological advantages, such as the avoidance of (consent) bias nor sufficiently address moral concerns such as the lack of attention to individuals’ autonomy and privacy. The latter argument also refers to the conditions of corporate data retrieval. Abandoning informed consent for big data research is seen as potentially hazardous to the reputation of academic research, in particular

with regards to public trust. Such arguments brought forward in response to big data-driven research indicate interdependencies between claims presented as part of different discursive domains: ‘strictly’ moral assumptions and the technological promises of big data can barely be treated separately from each other.

Validity claims to normative rightness (moral justice) as well as validity claims to truth (the factual accuracy of statements) need to be understood as co-constitutive in projects using biomedical big data for public health surveillance. Researchers particularly highlight societal benefits and future possibilities, from *normative* perspectives. They articulate claims to normative rightness, for example in terms of the desirability and expected benefits such as improved public health or cost effectiveness. But these claims to normative rightness are contingent on validity claims to truth, for example with regards to methodological conclusiveness and technological developments.

When considering the use of their data, individuals need to assess whether a certain claim to normative rightness, such as the safeguarding of privacy, may be seen as valid. Likewise, they need insights into the conditions and consequences proposed in related claims to truth: for instance, if the level of privacy proposed as morally reasonable can be indeed safeguarded by certain technologies and methodologies. It is therefore misleading to completely separate statements regarding a technology’s functional aspects from normative claims. Along these lines, Swierstra and Rip (2007, 7)⁹² even suggest that ultimately, all arguments brought forward in debates on new and emerging technologies are ethical.

In this sense, there is no difference between the ethical, legal, and social aspects (ELSA) in science and technology developments. Instead, ‘[p]resumably ‘non-ethical’ arguments in the end refer to stakeholders’ interests/rights and/or conceptions of the good life – thus, ethics’ (Swierstra and Rip 2007, 7). Swierstra and Rip stress that this notably applies to discourses on health and environmental risks, which are commonly, yet misleadingly, framed as mainly technological issues. In contrast, the authors emphasise links between technical and ethical matters, reasoning that ‘[...] the technical discussion can be opened up again to ethical discussion when the assumptions protecting the technical approach are questioned’ (ibid.). Bringing this back to Habermas’ emphasis on valid social norms as *just* norms, this means that in big data-driven health too, surveillance validity claims to truth and rightness *alike* amount to matters of social justice.

Therefore, to assess the moral reasoning of big data-driven research, we likewise require transparency in terms of methodological and technological conditions. The tech-methodological blackboxing, which is characteristic of big data-driven research, however, obstructs individuals’ possibilities to engage with validity claims to truth. The argument above also implies that realistic deliberations regarding big data’s contribution to public health are ultimately

ethical matters. A main reason for this is that articulated techno-social benefits are commonly mobilised to downplay concerns regarding civic, individual rights. These interdependencies are particularly relevant when considering the institutional conditions of big data-driven health research and its ethics.

Ethical (self-)assessment tends to be constructed as a 'protectionist hurdle': an obstacle to overcome, for example during the grant application process as well as at certain points throughout a study. Once a tech-oriented project has received the approval of the relevant Institutional/Ethics Review Board, or a comparable committee, there are few incentives to engage with ethical issues. For scientists involved in big data-driven research, continuous overtly critical, tech-methodological as well as ethical concerns are unlikely subjects to foreground. They are mostly incentivised to *justify* rather than question their innovation under competitive conditions.

Research projects commonly need to be presented in ways that enable scholars to acquire funding and to publish refereed papers. This leaves little leeway for stressing risks and uncertainties which could undermine a project's feasibility and competitiveness. In the context of big data-driven biomedical research, this has likely facilitated the tendency that contributions to the public good are commonly foregrounded, while ethico-methodological uncertainties are deemphasised. These dynamics also reflect more general insights into novel technosciences, as observed by Rip: 'Newly emerging sciences and technologies live on promises, on the anticipation of a bright future thanks to the emerging technology [...]' (2013, 196). In contrast, foregrounding ethical concerns may challenge the acceptance of innovations and undermine possibilities for funding in tech-centric grant schemes. This also raises the issue that funding programmes need to open up further possibilities for critical engagement with ethical issues.

Facilitated by the abovementioned factors, risks and ethical uncertainties tend to be deemphasised in comparison to benefits for the common good. Issues such as informed consent, privacy, anonymisation, research transparency and methodological sustainability, as well as entanglements between scholarly research and corporate data economies, are at best mentioned, but rarely scrutinised in ethical accounts of big data-driven research. With regards to privacy, scientists indicate that users' current legal rights and laws relevant to corporate data retrieval are decisive for their methodological choices. But critical research indicates that users' privacy *expectations* diverge from current possibilities for privacy management. Moreover, users' current rights and corporate responsibilities remain to be redefined in emerging legal frameworks and data protection policies.

By using, for example, social media data, researchers endorse their collection as morally reasonable. They foster the perception of such data retrieval as the undisputable status quo and the (future) way to go. This is especially problematic when considering the as yet meagre attention paid to potential ethical

issues concerning the role of internet and tech corporations. In a call for essays titled 'Fresh Territory for Bioethics: Silicon Valley', Gilbert (on behalf of The Hastings Center) observes that:

Biomedical researchers are increasingly looking to Silicon Valley for access to human subjects, and Silicon Valley is looking to biomedical researchers for new ventures. These relationships could be a boon to medicine, but they also raise questions about how well-informed the consent process is and how securely the privacy of the subjects' identity and data is kept. Other than a few quotes in the popular press, bioethicists have had little to say on the topic, although those whom I have spoken with agree that more attention is warranted. (2015)⁹³

Moral uncertainties and controversial issues, if at all, mainly appear as side-notes in big data-driven research. Those few researchers investigating ethical issues are often not directly involved in big data-driven research per se. This tendency speaks further to the juxtaposition of, rather than collaboration between, big data scientists and ethicists. Relating this back to the stakeholder constellations, this also means that there is not only little public discursive engagement: in addition, there is a lack of discursive interaction between scholars using big data for health research and those examining such approaches.

From Data-Driven to Data-Discursive Research

Ethical foresight has been emphasised as an indispensable feature of research involving new and emerging technologies (Floridi 2014; Brey, 2012; Einsiedel 2009). Grappling with ethical issues, risks and uncertainties should not be an approach taken in retrospect. Instead ethics should be an integral part of policy-making, regulatory decisions and developments (Floridi 2014, 501). It is characteristic for technological and scientific innovation, however, to move beyond the imaginaries developed in policy-making contexts. Before novel, ethical issues are negotiated in policy-making and governmental regulations, they may have unfolded in research or development phases already, as also implied in the pacing problem. This issue likewise applies to big data and their use in public health surveillance/research.

Therefore, ethical foresight should not be understood merely as a feature of regulatory practices (see also Swierstra and Rip 2007, 17). It is just as relevant to exploratory stages concerning new and emerging technologies, particularly with regards to their role in research. Ethical issues should be foregrounded and debated continuously, but they are often rather reluctantly taken up. Part of the issue is that the work of ethicists is often understood as the opposite of innovation. In contrast, a pragmatist approach to ethics emphasises that

moralties are likewise evolving in interaction with technological transformations, among other factors.

Given that valid social norms and ethics require formative discourse, we urgently need a shift from big data-driven to data-discursive approaches in research. What is currently neglected are inclusive, ethical debates on how the morals and norms pertinent to big data practices and particularly research are formed and justified: how are they developing and how should they develop? Whose positions are (not) reflected in these norms? This is also related to the more practical lack of consideration for how big data practices undermine prior modes of discursive involvement: is it ethically reasonable to abandon informed consent in certain studies and, if so, how can these studies provide novel ways to compensate for this?

From a discourse ethics perspective, this also means that research involving big data currently relies on norms whose validity is largely speculative with regards to the (dis-)approval of affected individuals. I therefore argue that researchers need to move away from big data-*driven* approaches, focused merely on techno-methodological innovation, towards data-*discursive* research foregrounding ethical controversies and risks as well as moral change. This discursive development needs to occur in combination with innovative approaches for engaging potentially affected individuals and stakeholders.

Wide, controversial negotiations of ethical decisions and moral principles are crucial for enhancing the validity of social norms. As already indicated above in relation to the conceptualisation of ethics as a field of innovation, such negotiations are considered to be constructive. Or, as Swierstra and Rip (2007) put it in emphasising the relevance of learning and discursive struggle: 'Since Machiavelli, political theorists have pointed out that struggle among an irreducible plurality of perspectives can be productive.' (19) When acknowledging the merit of struggle and controversy, the question arises how to encourage such dynamics and relevant debates.

First, a part of the answer lies in a point stressed above: ethical issues, risks, and contested moralities should not be downplayed, but foregrounded and made accessible to affected individuals in comprehensible ways. This demand of course invites criticism, as being utopian, not least because it conflicts with how academic funding and publication environments commonly function. Such a potential objection, though, highlights the relevance of research funding/grant schemes which do not treat ethical questions as a side-issue of emerging techno-sciences, but as core contributions and the path to innovation.

Second, the abovementioned question indicates the – of course already much debated – relevance of strategies for public engagement and participatory research approaches regarding new techno-scientific developments (see e.g. Pybus, Coté and Blanke 2015, 4; Moser 2014; Rowe and Frewer 2005; Wilsdon and Willis 2004). Within this domain, it also implies that there are certain kinds of debate and involvement which researchers should seek: with regards

to health research involving big data, particularly ethical controversies, risks, and changing moralities. The engagement of potentially affected individuals in formative discourses facilitates valid, just norms crucial to emerging forms of public health surveillance using big data.

Stakeholders' involvement and interaction amount to learning processes that have been described as productive struggle (Swierstra and Rip 2007). This emphasis on learning also points to the relevance of notions such as data literacy and (digital) information literacy. Such terms refer partly to the capacity of individuals to contextualise, process, and critically assess data and information which they encounter online (see e.g. Herzog 2015). According to Pybus, Coté and Blanke (2015), '[d]ata literacy can act as an extension and updating of traditional discourses around media literacy by refocusing our attention to the material conditions that surround a user's data within highly proprietary digitised environments' (4). However, they also point to the changing, precarious conditions under which researchers have come to access and handle big data (Haendel, Vasilevsky and Wirz 2012).

Data literacy is just as much a matter of technical expertise as of possibilities for discursive engagement and ethical debate. The importance of involving affected individuals also implies an understanding of data literacy as expertise and engagement which is distributed among multiple stakeholders. The above-mentioned lack of attention for contested moralities and norms in public health research involving big data highlights an urgent need for discussion of the ethical dimensions of data literacy. This applies to the ethical expertise invested in research projects as well as individuals' possibilities for realising, opposing or endorsing the use of their data on moral grounds. In this context, the concept of data literacy is not merely meant to imply users' capacities and responsibility to understand the employment of their data. Instead, it aims at stressing the need for an expertise in and sensibility towards issues beyond practicability and optimisation on the part of data collecting and utilising actors.

Data literacy is not simply a skill which corporations or researchers can demand from the public. Instead, they need to consider, and improve, how they play a part in its formation. Relevant knowledge and skills concerning the implications of new technologies, for example regarding the ramifications for individuals' autonomy, need to be acquired. For this process, public debate, controversy and struggle are crucial. As stakeholders in these debates and dynamics, potentially affected actors should not be simply seen as an obscure public that merely needs to be informed in order to be empowered. Instead, potentially affected individuals need fair chances and opportunities for realising and negotiating research practices which concern rights, risks, uncertainties and moral values. These negotiations may just as much result in approval as in disapproval of norms applicable to current big data. Yet this is a decision which needs to be worked towards by involving relevant stakeholders and creating possibilities for civic debate and engagement.

This demand stands in contrast to current tendencies in big data-driven studies that foster further alienation between researchers and those individuals generating data in the first place. With internet and tech corporations incentivising big data-driven research by offering data or funding, researchers need to account for interdependencies between corporate interests, research developments and ethics. To move towards valid social norms concerning the use of health-indicative big data, scholars need to treat and discuss these data not merely as a technologically enabled opportunity. Instead, they need to be foregrounded as matters of ethics and social justice.

