Deutscher Ethikrat

Big Data and Health – Data Sovereignty as the Shaping of Informational Freedom

OPINION · EXECUTIVE SUMMARY & RECOMMENDATIONS

30 November 2017

The full text of the Opinion as well as all publicly available information and documentation of the German Ethics Council accompanying the work on big data and health are available at http://www.ethikrat.org.

Published by the German Ethics Council

Jaegerstrasse 22/23 · D-10117 Berlin
Phone: +49/30/20370-242 · Fax: +49/30/20370-252
Email: kontakt@ethikrat.org
www.ethikrat.org

Title of the original German edition: Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung

English translation: Forrest Holmes
Layout: Torsten Kulick

>> CONTENTS

>> EXECUTIVE SUMMARY

Fundamentals: Big data and health

1) Big data is one of the key terms in present debates concerning the societal changes driven by technology. It refers to the processing of large quantities of data, with the aim of discerning patterns and thus gaining novel insights. The volume and variety of data, as well as the velocity with which it is captured, analysed and interlinked, requires the use of innovative and continuously evolving technological approaches.

2) Since the early modern period at the latest, the systematic collection and analysis of data has represented an important factor in civilisational development and has encompassed both humans and their environment, for example in biology and medicine, psychometrics, epidemiology and in the social sciences. The use of modern computers, data storage technologies and high-speed networks has brought about an enormous increase in the volume of available data. It has

also facilitated various qualitative advances, such as the use of ever more complex sets of processing instructions (algorithms) in processor-intensive computer simulations, and the rationalisation, standardisation and refinement of numerous work processes.

3) The development of big data entails a transformation of every stage of information processing, characterised by increasing automation and the interlinking and interpenetration of data. Both the volume and the speed of fully automated data collection have increased exponentially in just the past few years, and the rapid distribution and networking of a growing number of devices, capable of capturing data in every sphere of everyday life, continuously opens up new sources of data.

4) This is especially evident in the healthcare sector, where growing numbers of researchers, companies and doctors make use of information derived from the processing of immense quantities of data. Furthermore, the generation of health-relevant data by individuals, for example by means of smart phone apps and sensors worn on the body, is constantly increasing. The interlinking and analysis of this diverse data permits penetrating insights into an individual's state of health, personality and lifestyle, even allowing for predictions to be made regarding, for example, the development of an illness.

5) Once data is collected, it is exchanged and interlinked, often across state borders and occasionally in real time, by data networks and networked software systems. To achieve this, technical standards are being developed for the exchange of data via application programming interfaces (APIs). These also facilitate the establishment of data usage rules and the tracking of data.

6) The efficient collection, storage and processing of data requires powerful computers. Frequently offered by commercial providers, this capacity is usually supplied by data centres containing numerous

networked servers. The shift from local computers to the virtual space of data centres is referred to as cloud computing.

7) The objectivity, reliability, reproducibility and validity of the data and analysis methods employed are key to assessing any statements, conclusions or predictions based on this data. As the quantity of data increases, so too does the significance of the analysis for individual factors examined, as well as the ability to consider additional factors and their interactions, even those with only weak effects. Nonetheless, the independent review and verification of data analyses remains of central importance.

8) Statistical relationships between variables (correlations), cannot alone provide the basis for conclusions regarding causes (casual effects) or modes of action. These can only be elucidated through additional arguments and assumptions or by means of additional data, e.g. from long-term or experimental studies.

9) Machine learning is of particular importance for the use and further development of big data applications. Here, by means of training datasets, statistical models "learn" algorithms that allow data to be classified or categorised in certain ways. A central question this raises concerns the degree to which such techniques could lead to the emergence of machine agents with the ability and authority to make decisions, which could intervene, for example, in shaping therapy or healthcare policy.

10) Using data drawn from a large number of people, self-learning systems are able to discern important factors, such as health-relevant behaviours, and can locate individual persons and values within this system of coordinates. Such approaches can rapidly provide individualised recommendations and permit tailored interactions with machine assistants. However, they also necessarily entail the divulgence

of personal information and increase the potential for deception and the manipulation of personal decisions.

11) By analysing connections amongst disparate data points, processing methods that employ big data are able to recognise ever-finer distinctions between people. Thus highly personal characteristics and circumstances increasingly factor into decision-making processes – for example concerning medical diagnostics, prognoses and therapies, or in the insurance industry's assignment of policyholders to different premium categories. The use of complex big data algorithms to generate such groups (stratification) must, however, take account of and minimise any potential sources of error.

12) Health-related data that can be attributed to a specific individual is particularly sensitive, as it allow insights into an extremely intimate sphere. Such individual data can be captured and linked from a constantly growing number of sources. In the course of their analysis, even data that at first glance appears incidental, such as information pertaining to one's movements or shopping behaviour, can be revealed as relevant to the assessment of one's health.

13) Health-relevant data accumulates in various, partly overlapping contexts, from medical practice and health-related research to government agencies and insurance companies, including the active or unintentional generation of data by patients and citizens themselves. Furthermore, big data technologies facilitate the thorough de- and recontextualisation of data collected, analysed, and recombined for different purposes. This leads to a blurring of boundaries concerning what may be considered health-relevant; it also increases the likelihood that data could be de-anonymised, or individual persons be re-identified.

14) Because all data, regardless of the form in which it is generated, *could* be construed as being somehow related to personal health, it is in

principle possible to classify all such data as health-relevant. As a result of this development, it is often no longer possible to determine at the time of their collection whether certain data should be considered sensitive or health-relevant. Rather, this depends primarily on the context in which the data is used. This context may change over time.

15) Various actors with different functions and at least partially opposed interests, operating in a number of diverse contexts, are responsible for the collection, processing and use of massive volumes of data. Here, five areas in which big data is applied in relation to health can be identified and, treated as exemplary, examined in terms of potential benefits and risks: First, biomedical research; second, healthcare provision; third, the use of data by insurers and employers; fourth, the commercial exploitation of health-relevant data by globally operating IT and internet companies; and fifth, the collection of such data by affected individuals.

16) In biomedical research (area one) the analysis of large volumes of health-relevant data is meant to facilitate an improved understanding of scientifically important connections and processes. Among the most data-intensive applications are modern imaging and molecular biological procedures, such as those employed in the so-called 'omics' (e.g. genomics, proteomics, metabolomics).

17) Key actors in biomedical research include not only research institutes and their staff, but also research subjects and patients. The use of large volumes of data in research generally adheres to high and well-verifiable standards of data collection, use and security, and often involves multiple institutions. Scientific organisations avail themselves of the new technical and infrastructural possibilities of big data, networking with each other in order to exchange data and collaborate on its analysis and evaluation.

18) The interactions of factors that cause and modulate many diseases are extremely complex. Big data allows researchers to integrate and aggregate various information from across multiple sources in comprehensive analyses. Not only is the sheer volume of data employed essential to such an operation, but also the quality of their interpretation.

19) The amalgamation of data collected by various institutions in often very different contexts presents special challenges for the use of big data in medical research. Frequently, uniform standards for data collection, annotation and quality control are lacking; so too are well functioning rules for data exchange. This is due, on the one hand, to concerns over the protection of personal data and a dearth of suitable channels for communicating with, and models for obtaining the consent of, patients and research subjects regarding the secondary use of data. On the other hand, there exist uncertainty and divergent viewpoints regarding the question of who has the right to access research-generated data, and to what extent.

20) Alongside new models for obtaining patient or subject consent, potential solutions in this area include, in particular, technical measures for the standardisation of data exchange, which would guarantee data quality and high standards of privacy protection, as well as regulatory support and initiatives to promote open data exchange.

21) In the area of healthcare provision (area two), the use of big data presents opportunities for the development of more highly personalised treatment plans as well as for the improvement of efficacy and efficiency. The drawing upon large volumes of data can refine the stratification of patients, in order to, for example, mitigate side effects and avoid unnecessary therapies. The collection and evaluation of health-related data also opens up new possibilities for the early detection and prevention of disease.

22) The healthcare sector is characterised by a multitude of actors with partially divergent interests. They include healthcare providers, insurers and patients, as well as governments, interest groups and researchers who are directly involved in clinical practice.

23) Alongside the opportunities presented by data-intensive approaches, there are also risks for patients, who stand to forfeit control over their own data, and who face the ever deepening penetration by healthcare suppliers into their most intimate sphere ("the transparent patient"), alongside the heightened risk of data abuse. To these risks can be added concerns that the increasing employment of healthcare approaches based on big data could further reduce the personal attention medical professionals devote to their patients, and that their uncritical or improper application could lead to errors in diagnosis and treatment.

24) For employers and insurers (area three), big data offers wide-ranging possibilities for accessing and analysing information of value, an area which current legal provisions do not always sufficiently cover. Ever-expanding volumes of data, and new ways to link this data, allow for increasingly fine-grained profiles of individuals or groups of people.

25) This has given rise to concerns regarding the potential for discrimination, in view of plausible scenarios in which insurers and employers, by analysing commercially available personal behavioural profiles generated using big data, could selectively choose only low-risk applicants or candidates, or offer them better conditions.

26) Even with existing contracts, employers and insurers have a stake in the health of their employees and policyholders, as illness can generate major costs. The monitoring of patient or employee behaviour allows for the introduction of incentives to encourage a healthy lifestyle, or sanctions to discourage an unhealthy one. Insofar as such programs result in the reduction of illness, they offer attractive prospects for

everyone involved. However, the risks cannot be ignored. Neither the adjustment of one's insurance premiums, nor disciplinary warnings received for behaviour detrimental to one's health, are in the interest of those who share their health data.

27) Global IT and internet companies (area four) primarily take the form of service providers. On the basis of their access to enormous quantities of data and their command of the necessary data infrastructure, they are able to provide search engines, interactive information platforms and offers such as online shopping, but also a broad range of multifunctional devices. Multitudinous user data is thus gathered on a vast scale, stored and exploited. For such companies, which are increasingly active in areas relevant to healthcare, it is thus uniquely possible to link data pertaining primarily to health with numerous other kinds of information. This implies a major potential for misuse.

28) Companies offer software, hardware, technology development, and online services for big data applications. They provide data-oriented institutions with access to systems, algorithms, devices and infrastructure for data collection, analysis, management and storage; the aim is to accelerate and improve processes and to ensure highly efficient use of relevant information.

29) The increasing activity of digital firms in the healthcare sector presents opportunities for research and medicine: compared to the public sector, major internet companies have access to considerably larger volumes of data, and are often equipped with better technical and financial resources as well as more powerful means of data analysis. On the other hand, by restricting data access for those who originally provided the data, or for those interested in using this data for medical or research purposes, private firms can also potentially impede medical progress.

30) The collection of health-relevant data by affected individuals themselves (area five) is facilitated by numerous wearable devices featuring sensors and apps, with which ever more data pertaining to an individual's health, activities, and environment is captured, processed and combined with existing data stocks. Moreover, the digitisation of everyday life has advanced to the point at which ordinary behaviours and forms of communication automatically entail the generation of data – frequently even beyond the scope of social networks, lifestyle apps and similar services.

31) Devices and apps that collect health-relevant data can make it easier for their users to access their own health information regardless of time or place, and can facilitate the provision of evidence-based healthcare. They can also promote a health-conscious lifestyle and further one's personal wellbeing. Additionally, they offer the possibility of enhancing research when used as an important quantitative and qualitative supplement to existing bodies of data.

32) On the other hand, an excessive regime of self-control aided by such services and devices can contribute to an exaggerated drive for optimisation detrimental to personal health, as well as the medicalisation of 'natural' life processes. Furthermore, the question arises as to whether self-tracking indeed is an expression of personal sovereignty, or whether it instead represents a form of self-induced heteronomy. Also of concern is the potential for discrimination against persons unable or unwilling to subject themselves to such measurements. The fact that many of the self-tracking apps and devices presently available are oriented towards the economic interests of their manufacturers, alongside an inadequate user-friendliness, transparency and privacy protection that many exhibit, has also been cause for criticism.

33) In summary, the following strengths, weaknesses, opportunities and risks can be identified in relation to the growing presence of big data in health relevant areas, irrespective of the context of application: The

strengths include the growing size of databases and the associated development of innovative digital instruments, as well as the high level of networking by the actors involved. Among the weaknesses are inconsistencies in data quality, a lack of transparency of data flows and loss of control over data, as well as increased demands for control, regulation and qualification.

34) The opportunities posed by big data consist above all in improved possibilities for stratification in diagnosis, therapy and prevention, the resulting improvements in efficiency and efficacy, and the encouragement of health-promoting behaviour. Risks are posed by the erosion of principles and practices of social solidarity, the diffusion of responsibility, monopolisation, data misuse, and informational self-endangerment.

35) How specific health-relevant big data applications are to be judged, however, depends to a key extent on the actors involved, their various interests, their own assessments of risk and opportunity, as well as the particular context of application.

Legal provisions regarding big data

36) Big data represents a significant challenge for the legal system. Of particular relevance in this context are Constitutional law, general data protection legislation and special data protection provisions pertaining to the healthcare sector, and medical device regulations, but also underlying incentive mechanisms and self-regulative and hybrid steering mechanisms.

37) The key elements of data protection law are constituted at the level of the German Basic Law. The core constitutional standard at the national level is the right to informational self-determination, a principle elaborated by the Federal Constitutional Court, in a landmark

judgement on census participation, as a specific instantiation of the general right of personality. It buttresses and extends the constitutional protection of privacy and freedom of conduct.

38) The constitutional right to freely develop one's own person can collide with matters of concern to the common good, such as promoting scientific progress or ensuring effective and adequate healthcare. Conflicts can also arise with the fundamental rights of other private subjects who want to access and utilise data available to them.

39) Data protection law orients itself towards these constitutional provisions. Nonetheless, it is applied in many contexts that have only come about as a result of new technical developments, and for which it was not originally designed. Even the most recent amendments adopted in line with the European General Data Protection Regulation (GDPR) have not rendered it sufficiently adapted to the phenomenon of big data. This applies despite the clear progress that these new provisions represent in terms of, for example, the establishment of cross-border standards or the stronger attention paid to the concept of privacy by design.

40) The underlying assumptions, central principles and objectives of traditional data protection law can hardly be reconciled with the unique characteristics of big data applications. The basic principles of traditional data protection law – definitions regarding the personal nature of data, acceptable uses for data and the obligation to adhere to these, the necessity, proportionality and minimisation of data collection, the need for consent and transparency – stand in opposition to the particular logic of big data. If we do not simply wish to issue a general prohibition on the use of big data, while at the same time refusing to accept the significant curtailments in protection it entails, then we must devise new forms of regulation and ways to shape developments in this area.

41) As it stands, data protection law concerns itself with the personal nature of data and places particular emphasis on the need for its use to remain specific to an intended and explicit purpose. It is in the nature of big data, however, that the future use to which data may be put is not foreseeable at the time of its capture, and that the connection between data and a person or their health is only, at least under certain circumstances, established ex post facto. Data that has been stored for one purpose is often analysed in connection with another, or data is simply collected for as yet undetermined purposes.

42) Furthermore, big data is obviously incompatible with the principle of data economy or minimisation, according to which as little personal data as possible is to be collected, processed or used. If fully applied, this principle would easily lead to a far-reaching nullification of the possibilities presented by big data. Because the potential danger posed to the right to informational self-determination increases in proportion to the volume of data stored, however, more effective data protection mechanisms are needed.

43) Big data also reveals itself to be incompatible with the obligation to obtain consent, as governed by current data protection law, whereby data use is only permitted when consented to by the persons affected in full view of the nature and extent of the intended data use. Even as it stands, there is considerable reason to doubt that persons supplying data are fully cognisant of the uses to which their data will be put and the implications thereof. Big data significantly compounds this general problem, as the future uses of data are often simply unknown at the time these data are collected.

44) Moreover, beyond the point of consent, existing data protection law offers only few possibilities for influencing the subsequent fate of data. Every further use requires its own granting of consent, and once data is collected with consent, it can no longer be tracked by those affected. The dynamic of big data does not fit within this regulatory

model. Especially if one regards the consent of affected persons as a central requirement of data protection, new avenues must be explored by which this would be possible and functional under the conditions of big data.

45) Additionally, through the combination and interlinking of diverse data, big data increases the chances of re-identification and undermines the effectiveness of anonymisation and pseudonymisation requirements. To what extent and at what point is the danger of the re-identification of anonymised data, taken in isolation, sufficient to warrant an assumption that this data is subject to protection as being of a personal nature? This question only adds to the problems surrounding the already contested concept of personal data in data protection law.

46) The right to be informed about the collection of personal data, as well as the right to its correction, deletion, and blocking, serves the purpose of transparency, but often provides little effective protection. Especially in the context of big data, persons supplying data will hardly be able to identify all potential parties against whom claims can be asserted. Furthermore, the requirement that the ways in which data is processed be comprehensible to those who have provided it, which is encompassed by the right to information, proves difficult to realise in light of the complex and self-learning algorithms used by big data. Thus the rights to correction and deletion are nullified, as affected persons cannot avail themselves of these rights without first being fully informed as regards their data.

47) This analysis of the deficits of general data protection law can also be applied, with certain qualifications, to the special area of health data protection law. The latter supplements data protection law, which is often adapted to specific areas of application, with the civil, criminal and professional legal provisions pertaining to patient confidentiality. Ultimately, however, the possible solutions offered by health data

protection law also remain to a large extent trapped in an understanding of the problem predating the advent of big data.

48) The provisions of medical device legislation, which aim to regulate the free trade in medical devices while guaranteeing the safety, suitability and performance of such devices for the protection of patients, users and third parties, could have a compensatory effect. Unlike medication, medical devices do not require government approval, but must nonetheless be certified in accordance with a product-specific risk assessment, risk-minimisation and risk/benefit analysis, as well as a conformity assessment tailored to the risks inherent to the product.

49) Software that serves medical purposes can be classified as a medical product. Whether this is the case depends to a key extent on information provided by the manufacturer. In practice, however, the distinction between medical applications and mere lifestyle or fitness apps is often difficult to make.

50) The provisions of health insurance law also prove relevant with regard to big data. By including the costs of M-Health applications in their coverage, both private and statutory health insurers could, for example, generate financial incentives for the developers of these products, as well as offering an alternative to a "pay with data" model. It remains, however, to be demonstrated whether this would be an effective approach. Furthermore, the danger of discrimination would need to be avoided, including with regards to the use of such data in determining insurance premiums.

51) In light of the recent, thorough reform of data protection law through the GDPR and the new version of the *Bundesdatenschutzgesetz* (Federal Data Protection Act), it remains to be seen if and how new regulations and mechanisms in this area prove to operate. Nonetheless, it is clear that many of the basic principles of current data protection law remain hardly reconcilable with the concept of big data. Flexible

regulations, open to innovation, and operating within the leeway granted by constitutional law, can take this inherent tension into account, alongside the potential use of complex, civil-law and cooperative civil-state regulatory inputs.

52) Of particular importance would be to ascertain the extent to which the lack of concreteness that characterises health-relevant big data applications could be compensated for by additional technical and organisational, as well as material and procedural, safeguards. As data protection law continues to evolve, it is above all a more fully differentiated model of consent, giving room to the particular features of the regulatory domain and the preferences of those affected, or a strengthening of the collection and use of data on the basis of legal sanctions, that come into view in this regard. Civil law will also play a major role in the evolution of data protection, especially consumer law, liability law, and regulations pertaining to the assignment of data possession and the authority to determine its use (data 'ownership').

53) All regulatory approaches to big data must confront the problem of responding to an inherently global phenomenon with the legal apparatus of a territorially bounded state. Existing data protection laws, viewed internationally, vary widely, presenting both those affected by big data and those who seek to regulate it with unique challenges. Despite numerous efforts to harmonise data protection measures, numerous practical obstacles continue to stand in the way of an effective cross-border application of the law.

54) In light of the specific dynamics and volatility of this regulatory domain, cooperative solutions developed outside of state authority gain in importance, such as the certification of products with data protection or data security seals, or the development of codes of conduct or best practice in science and the private sector.

The ethics of big data and health

55) Big data impacts both ethical frameworks that are normatively and descriptively concerned with the role, function and position of the data-providing individual, and key axes of social orientation. The relevant concepts include freedom and self-determination, privacy and intimacy, sovereignty and power, beneficence and non-maleficence, as well as justice, solidarity and responsibility.

56) The concept of freedom is used in a number of different ways. A distinction can be drawn between authorship of one's actions as a basic condition of freedom, on the one hand, and self-determination as the practical instantiation of freedom, dependent upon more or less clearly perceivable circumstances, on the other hand. The authors of actions can be self-determined to varying degrees.

57) The concept of self-determination refers to the ability of a person to shape their life in accordance with their own ideas, as well as the actual putting into practice of this ability and an ideal way of leading one's life. These forms of personal self-determination must be differentiated from the legal protections afforded their exercise. The ways in which self-determination can be exercised, and the degrees to which this takes place, are of considerable practical importance. Thus one can, in certain contexts, delegate one's right to self-determination, or partly compensate for restrictions to one's capacity for self-determination through representatives.

58) In the context of big data, the last years have seen the development especially for biobanks of new models for acquiring consent, which, with a view to the self-determination of data providers, seek to strike a balance between an unrealistically narrow scope and an excessively broad authorisation of data use. Here, dynamic models where consent is repeatedly sought with regard to individual elements of data use are complemented by further options, such as possibilities for

delegation. Participants can furthermore decide as to which form of consent they fundamentally prefer.

59) In order to evaluate self-determination, we must also take into account the social context in which an actor is embedded. To be free and to be able to act with self-determination means, in this light, at least the realistic possibility of preserving and shaping one's identity, while taking responsibility towards oneself and towards others for one's actions. This necessitates reliable and fair standards under the rule of law, which apply equally to everybody.

60) Privacy is classically defined as the right to be let alone, or, in other words, as a sphere of personal existence from which the need to justify oneself or submit to unwanted public scrutiny have been largely excluded. Closely associated with privacy is the concept of intimacy, which defines areas of one's life reserved only for those immediately involved, and of which any details are made available to selected third parties only with express consent, if at all.

61) To a considerable extent, ideas about what is deemed private or intimate are culturally variable. This aside, the preservation of the private sphere can be normatively justified on the basis of its major social anthropological significance. Only in a private sphere can close social relationships and the conditions for personal development be formed. Privacy creates a space for intimacy and familiarity, in which people can attend to relationships and, without disguise or inhibition, truly be themselves – shielded from the outside, but open on the inside.

62) With regards to big data, potential threats to privacy arise from the numerous, novel opportunities presented for collecting, analysing and recombining data and information, as well as from the concomitant, increased difficulties in ensuring anonymisation and pseudonymisation. The more intimate details can be surrendered digitally, the

higher the risk of self-induced external control or informational self-endangerment, in the form of a personal lifestyle that renders itself significantly dependent on external influences.

63) Even if total control over one's data trail is impossible in a digital society, people nonetheless consider it important that they be able to determine, dependent on the given context, how their data are used and reused. At the same time, data users are increasingly expected to handle the data available to them in a confidential and trustworthy manner, even when decontextualising and recontextualising this data.

64) The question of how to protect privacy under the conditions of big data impacts not only the individual, but also groups. The analysis of large volumes of data often reveals combinations of characteristics shared by numerous people. Those affected are grouped by algorithms, with potentially stigmatising, discriminating or exclusionary consequences. Individuals are often unaware of being categorised in such a manner.

65) Of central significance in the context of big data is the concept of sovereignty. With its cultural-historical origins lying predominantly in the politico-religious domain, the notion of sovereignty takes on various concrete forms across numerous areas of life. It was understood as that property of God or of an absolutist ruler by virtue of which he, fully and without regard for other powers, could do or allow anything. In place of this supposedly absolute freedom of the sovereign subject, other understandings of sovereignty emphasise the ways in which a subject's physical and social embodiment are dependent on outside factors.

66) According to an understanding of sovereignty that at least fundamentally excludes the notion that one person may posses power over another, data of a personal nature are merely lent to their collectors

and users; these data are not freely and arbitrarily available property. Conversely, however, this does not mean that data providers are automatically the owners of their data, nor does it mean that they can realise their claim to sovereignty under all circumstances. Nonetheless, this notion entails broad capacity for control on the part of the individual.

67) The concept of sovereignty is closely linked to the concept of power. Sovereignty is realised in the exercise of power; conversely, it is limited by others exercising their own sovereign power. In the context of big data, specific ways of exercising power are of ethical significance: Firstly those by which the preferences and convictions of others can be manipulated; and secondly those that go further, even allowing a subtle shaping, modification and therefore a potential governing of others' characters.

68) The use of big-data algorithms provides those offering online services with new possibilities for exerting targeted influence on the thoughts, feelings, and actions of their users. The spectrum reaches from open nudging, by which health-conducive behaviour is to be subtly encouraged, to covert and manipulative interventions that, crucially, are designed to the benefit of others. The latter are, at the least, urgently wanting of ethical justification, for they evade the cognitive control of those they target, circumventing the affected individual's ability to govern the conditions of his or her actions and thus undermining their self-determination.

69) Another relevant normative point of reference emerges from the moral obligation of beneficence, according to which one's actions in numerous situations should be weighed such that they result not merely in the minimisation of cost, but should also benefit others, in particular those in need. Two aspects of the notion of beneficence are of particular interest in connection with the topic of big data and health: The expansion of knowledge and understanding, as well as

improvements in therapy, resulting, for various involved parties, from the new possibilities opened up by digital information gathering and the processing of large volumes of data in the health care sector.

70) Knowledge and insight are of major importance in the self-constitution of the individual and his or her ability to live autonomously. Furthermore, the critical examination, safeguarding and expansion of bodies of knowledge fulfils an important societal function.

71) The safeguarding of communication bound to truthfulness is necessary in order to achieve the goals associated with the advancement of knowledge. Particularly in the sciences, sophisticated methodological and theoretical standards have been developed in order to ensure such communication. Therefore, care must be taken not to allow new digital methods of data capture, analysis and recombination to lead to a relaxing of epistemological standards or to a loss in the reliability of the evidence they generate.

72) Questions also remain as to which groups should primarily benefit from the advancements in knowledge made possible by big data, how existing obstacles to a more efficient design of the data use process can be overcome, and how a just distribution of the positive effects that result from the anticipated gains in knowledge can be achieved.

73) The collection and transmission of large volumes of health-relevant data touches on fundamental questions of justice. As a normalising principle of social relations, justice demands that the arbitrary privileging of certain persons or groups be avoided. Rather, it is to be determined on a rational basis what is fair and proportional for every individual. This requires that uniform criteria be applied and differences in the treatment of various persons be normatively justified in a manner capable of achieving social consensus.

74) As regards big data applications in the healthcare sector, four sets of problems stand out as especially relevant to questions of justice: first, access to datasets for the research sector; second, the insidious consolidation of monopolistic structures; third, the inclusion of health apps, as well as various devices that facilitate private self-tracking, in determining health insurance premiums; and fourth, aspects of social justice, understood in terms of the capabilities approach, as they concern the responsible handling of health-relevant data.

75) The concept of solidarity denotes prosocial behaviours, practices and dispositions, as well as institutional, political and contractual regulations, the purpose of which is to assist others. Solidarity is frequently understood as complimentary to – and often subsidiary to – the concept of justice. It regularly emerges against the background of a group's common goals, in the face of a common challenge or from a shared idea of the good life within a mutually supportive community.

76) Solidarity is frequently grounded in expectations of reciprocity. The willingness to act in solidarity can diminish when doubts arise as to the realisability of such expectations. This can occur, for example, when in the long run the impression takes hold that others' need for help and support is inflicted upon themselves as a result of their own negligent behaviour and lack of initiative, thus overstraining the fabric of solidarity.

77) The ability, granted by big data, to analyse ever more comprehensive and diverse health-relevant data allows for the generation of more precise risk profiles. The concern hereby arises that the basic assumption of solidarity upon which the statutory health insurance system, as well as the fair structuring of contracts in the private health insurance industry, rest – namely, that vulnerability to health risks is something shared by all – could be called into question. This would allow low-risk groups to abandon in greater numbers the mutually

supportive group of statutorily insured people, placing substantially increased burdens on those who must remain.

78) Within the system of statutory health insurance, premiums set on the basis of behavioural data undermine the notion of solidarity that calls for protection against illness-related vulnerability largely without regard to risks deriving from individual behaviour. Private health insurance, on the other hand, operates with risk-based premiums. Here too, however, a redistribution of risks to the advantage of policyholders could result if, even after the conclusion of an insurance contract, future premiums were regularly adjusted on the basis of the continuous collection and analysis of individual data made possible by big data. This would wholly negate a core principle of insurance coverage, by which risks are mutually borne by a large group and premiums cannot be individually tailored. The potential would grow for smaller pools of policyholders to form, where cases of illness or injury would more quickly lead to increased premiums.

79) Furthermore, private insurance policyholders not willing or able to participate in a behaviour-based insurance model could be denied financial incentives; over the long term, this would lead to disadvantageous premiums. Regardless of whether or not they pursue a healthy lifestyle, these policyholders would be punished for not granting their insurer access their personal data, and would thus be placed at a disadvantage simply for exercising their right to informational self-determination.

80) Fundamentally, the freedom to live life, and develop one's personality, according to one's own design has priority over a strict and permanent obligation to avoid all health risks. While this principle does not apply under all circumstances, it does render it difficult to qualify a perpetual, targeted collection of data on one's individual lifestyle, or the use of risk profiles, fed by big data, encompassing all areas of life,

as a reasonable expectation of responsibility to which one could be held for one's own health.

81) Whether and how statutory health insurers could take account of policyholders' personal responsibility, and influence their health-related behaviour, remains a topic for debate. Data-derived incentive structures could develop into highly intensive and invasive forms of surveillance. On the other hand, a sophisticated uncovering of risk factors using big data analysis, integrating data from all areas of life, could in the future reveal that the overwhelming majority of the population is characterised by mixed risk profiles, encompassing factors both favourable and negative, and of a physical, mental, behavioural or other kind.

82) In various areas of medicine, the application of big data technologies has already led to the development of new, prosocial mutual support practices such as, for example, the formation of small groups of patients who share the same experiences or risks of rare diseases. This allows them to combine their data and biological samples in collective pools, to be placed at the disposal of research into their particular sets of symptoms.

83) Other advances in terms of solidarity can currently be seen in online forums, where patients can input, exchange, discuss, and make use of for their own health management both self-collected and clinical information and experiences. As the development of online, networked self-help instruments for patients accelerates, the expansion of such practices is to be expected.

84) As a moral category, responsibility can be differentiated according to types of action and decisions as well as according to the arrangement of institutional structures. Responsibility can be demanded and taken up morally, legally, politically and contractually both before and after an action or decision. The various corresponding sorts of

responsibility often exist in an objectively reciprocal relationship: one expects the assumption of responsibility for the future from precisely those parties that one would call to account in an actual case of injury. The complex interplay between individuals, institutions and technology entailed by the use of big data takes on a particular importance in areas relevant to health and healthcare. What must be avoided is an opaque diffusion of responsibility, which constitutes a danger in any situation involving the interaction of numerous actors and highly technical processes.

85) Particularly in the big data era, a certain framework is required to enable individual data providers to take responsibility for their data. This framework should be technically and organisationally effective and easy to use. In the especially sensitive domain of health and healthcare, furthermore, a heightened duty of care pertains, for example, for researchers or doctors.

86) Key among the ways for businesses to engineer responsibility into their big data processes is the need to create the fundamental conditions for responsible data management, to render already granted consent revocable, and to design easily accessible options for data administration. One could exempt from these requirements sufficiently aggregated data, derived data, or cases that demonstrably preclude the identification of individuals from the data. To use such approaches to facilitate the de- and recontextualisation processes specific to big data, while simultaneously safeguarding high standards of anonymisation, and to create confidence in institutions using big data, is a key task that lies ahead.

87) Another way for industry to take responsibility for the rights of the individual while also protecting legitimate business interests would be to use application programming interfaces to set up delegation systems. Such interfaces could act as "data agents" in implementing data providers' preferences regarding the handling of their data. Thus,

individual data management would be replaced with a programmatic data management system, granting individuals a reliable and technically accessible means of assuming responsibility for choosing short-, medium-, and long-term strategies for the handling of their data, while eliminating the need to reach a separate decision on every question of data use.

88) Businesses could also assume responsibility by strengthening the oversight and verifiability of their processes in terms of, for example, the algorithms employed; the measures taken to eliminate systematic discrimination; the adherence to regulations pertaining to data safekeeping, anonymisation and deletion; and the gapless and tamperproof of the origin, processing, use and exchange of data.

89) Apart from governmental regulation, there exist other ways to ensure and/or promote the assumption of responsibility by institutional actors. Certification, seals of quality, or voluntary standards established and overseen by interest or industry groups could, for example, strengthen confidence in the organisations and processes concerned.

90) Another question of responsibility concerns organisations potentially encroaching upon the personal communication between users in the form of, for example, tips and offers promoting healthy lifestyles. On the one hand, the objection to obvious intrusions into the private or intimate sphere would speak against such actions. On the other hand, if the functional reliability of the underlying algorithms were scientifically well substantiated, one would be compelled from an ethical perspective to take into account the possibility that such actions could prevent major suffering or even death, as in, for example, offers of assistance in social networks targeting persons at risk for suicide.

91) The state can assume responsibility on the national level, as part of the EU, or as a party to international law. With a view of the abovementioned difficulties surrounding legal enforcement, however,

a principle of regulatory subsidiarity should prevail, by which voluntary obligations and certification take priority over detailed legal regulations, provided the former function effectively.

92) In terms of the three levels of potential allocation of responsibility in the field of health-relevant big data applications (individuals, organisations, and the state), individuals are indeed obliged to assume responsibility for the use of their data. Nonetheless, it is primarily the duty of those organisations collecting, processing and passing on these data to ensure the conditions for the responsible shaping of informational freedom on the part of the data provider.

93) The less organisations are willing or able to make available the technical means by which individuals can more easily control their data, the more pressing the need for the state, from the perspective of an ethics of responsibility, to intervene in order to guarantee, oversee and, where applicable, regulate and sanction. The goal of giving the individual the capacity for a sovereign relationship with their data is only attainable when the requisite responsibility is taken up on all sides.

Data sovereignty as the shaping of informational freedom

94) Data sovereignty, understood as the responsible shaping of informational freedom, in a manner appropriate to the risks and opportunities presented by big data, is the central ethical and legal goal in confronting the challenges and opportunities presented by big data.

95) The notion of shaping informational freedom builds on the concept of informational self-determination. It is not grounded in exclusive rights analogous to property, but rather in each person's authority to determine with which content one chooses to relate to the wider world. Shaping informational freedom in this sense refers to the

interactive development of one's personality in a networked world, and is characterised by the capacity to intervene effectively in the constant flow of individually relevant data on the basis personal preferences. The shaping of such freedom is responsible when it also orients itself towards the legal and societal demands of solidarity and justice.

96) With the concept of data sovereignty as described here, we seek neither to perpetuate the established, barely modified regulatory approach of data protection, nor do we call for a total reorientation, let alone the abandonment of the conventional notion of data protection, or the general lowering of the existing level of protection. Rather, the aim is to fulfil and render effective the basic normative requirements described above, including those pertaining to an informational self-determination grounded ethically and in terms of basic rights, and thus pertaining to data protection, under the novel conditions of big data.

97) Data protection is not, nor has it ever been, an end unto itself. Rather, it serves to protect the person – both their private sphere as well as the free development of their personality in the public sphere. With the concept of data sovereignty, however, we also wish to emphasize the aim of combining the individual's sovereign, i.e. self-determined and responsible, handling of their own personal data with a realisation of the potential opened up by big data, both for society and for shaping individual lives.

98) As a goal, the responsible shaping of informational freedom in the domain of health and healthcare consists of taking full advantage of the potential opened up by big data for medical research, clinical practice and individual health and health-related behaviour, while reducing the concomitant risks to a minimum.

99) In terms of shaping and exercising data sovereignty, two increasingly proximate and, in places, already overlapping spheres can be distinguished: First, there is data use in medical research and clinical practice, which has thus far been characterised by relatively clear and stringent data protection, quality, and security standards; second, there are the extremely heterogeneous products and services on the free market that increasingly exercise influence over developments in the healthcare sector. The latter category extends from application concepts that border on the first sphere and its standards to evidently dubious products uninvested in the sustainable promotion of health.

100) Developments in the field of big data cannot be stopped, but they can certainly be steered. Because the mechanisms and forms of action that characterise traditional data protection law do not suffice to do so, the task ahead is to devise models for regulating and shaping these developments that more faithfully reflect their complex and dynamic nature. They should reflect the principle of data sovereignty as the shaping of informational freedom in a multidimensional way and with view to various groups of actors and contexts of action, taking up the potential forms and attributions of responsibility described above.

101) Under big data conditions, it is necessary to abandon outdated notions of a particular kind of data having a specific, given sensitivity, that evokes corresponding protective mechanisms. Data protection can no longer be statically tethered to certain categories of data and data use; rather, it must adapt to the constant recombination and recontextualisation of data.

102) A model designed to regulate and shape data use that orients itself toward the principle of data sovereignty focuses on those persons who provide data as the key actors to be protected and respected. The goal is to empower these subjects, as well as the organisations connected to them, to achieve sovereign handling of their data, by

designing regulations and shaping institutions in a manner sensitive to context and appropriate to each case. Simplified, wholesale solutions should be avoided in favour of more complex, institutionally diversified, compound models that are flexible and appropriate to the problems posed.

103) The heterogeneous second sphere described above should be shaped according to the following core principle: the more closely an individual application borders on the first sphere, the stronger the ethical and legal imperative to steer its development, with reference to multiple actors, towards the standards of quality, protection and confidence that generally pertain in the first sphere.

>> RECOMMENDATIONS

The German Ethics Council recommends a governance concept oriented towards the central goal of data sovereignty. Such a concept calls for a comprehensive societal effort including both legal and extralegal elements, and incorporating technical advancements made available to all societal actors in such a way that guarantees the preservation of basic rights. The governance concept presented here contains specific recommendations for action in four areas. These aim to, firstly, realise the potentials of big data; secondly, to ensure individual freedom and privacy; thirdly, to ensure justice and solidarity; and fourthly, to promote responsibility and trust. The measures recommended here should be financed and implemented as soon as possible.

A. Realise potentials

In order to realise the potential benefits of big data in the healthcare sector, the cooperation between numerous actors from clinical practice, basic

medical research, and companies involved in health-related fields, as well as individual data providers must be as seamless as possible. The goal should not only be the prospective collection of and sustainable access to datasets, but also to facilitate the combination of already existing datasets from the clinic and from research with newly acquired data in an ethically responsible manner.

A1. Create the necessary basic infrastructure

Being able to take advantage of the potential of big data in the healthcare sector vitally depends on the availability of a high-performance infrastructure for gathering, storing, analysing and transmitting large volumes of data. In order to avoid problematic dependencies on commercial providers for these infrastructural services, who are often not subject to German or European data protection standards, public authorities should ensure that such infrastructure – especially for clinical practice and basic medical research – is built and developed promptly, that adequate access is facilitated and that it is subject to public oversight.

A2. Facilitate data exchange and integration

It is just as important that the responsible exchange and integration of health-relevant data between multiple institutional actors is ensured by a number of measures and the sufficient public funding to implement them:

A2.1. Develop and make available standardised data interoperability procedures

In order to enable an adequate aggregation of data from different sources, whilst respecting data providers' right to privacy, data must be comparable with other data; i.e. it must be consistently labelled and appropriately annotated. Crucially, this requires the standardisation of data formats and the creation of quality control options, including transparent documentation of the steps taken.

A2.2. Refine cooperative research data management
Current initiatives to establish structures for efficient communication, collaboration and coordination between participating institutions should be bundled, intensified and given a long-term perspective. At the same time, attention must be paid to ensuring adequate interfaces with the telematics infrastructure, as well as alignment with the further development of data exchange within the healthcare sector, as specified in the *E-Health-Gesetz* (E-Health Act).

A3. Promote and protect data and research quality
A key task moving forward is ensuring data quality in order to produce sufficiently reliable results. The following measures are necessary to achieve this:

A3.1. Adhere to epistemic standards, especially those of evidence-based medicine
As mechanisms for controlling the safety and efficacy of medical interventions are further developed such that they can deal with big-data applications, the established standards of evidence-based medicine must not be undermined. If serving medical applications, processes based on big data must also be subject to established clinical tests for safety and efficacy.

A3.2. Introduce uniform data and documentation standards
The introduction of uniform standards represents a sensible measure, not only in terms of facilitating interoperability and cooperation, but also in order to ensure effective quality control. For example, this includes questions regarding data and metadata formats, the step-by-step reconstruction of the data-use process, version control, and the mapping of semantic links and hierarchies of data in the most consistent way possible. In particular, quality-assurance standards for data should include documentation requirements to facilitate tracing the origins of data and, at the very least, their future traceability.

A3.3. Establish data quality seals
In order to render transparent the aforementioned quality standards and their underlying requirements, compliance certificates ("quality seals") should be awarded that verifiably demonstrate the origin and quality of the original data and the processing steps they have undergone (e.g. by means of blockchain technology). Because quality assurance is also in the various actors' own interests, the primary focus should be on internal monitoring mechanisms in science and industry. Insofar as these prove to be deficient, however, overarching legal requirements should also be introduced.

A4. Adapt the legal framework for data use for research purposes

A4.1. Further develop the secondary use of research data
Where applicable data protection law allows, based on a careful weighing of interests, the processing of personal data even without consent – if data serves and is indispensable for scientific, historical or statistical purposes (Section 27 of the Federal Data Protection Act, new version) – then additional procedural protection and design measures such as cascading consent models (see recommendation B2) should in principle be employed, in the interest of data sovereignty.

A4.2. Facilitate the individual's legal options to allow the full use of their data for medical research purposes ("data donation")
In principle, the traditional consent model requires that personal data only be collected within strict limits prescribing their intended use. Precisely because the consent model must be adhered to, not only should its procedures be broadened, but they should also become more open for specific domains. Specifically, this should facilitate the individual's ability to allow, by means of a comprehensive consent agreement, the use of their data, without strict earmarking, for the purposes of basic clinical and medical research ("data donation"). The prerequisite would be a full clarification of the possible consequences, particularly with regard to the rights of others, such as affected family members. Also necessary would be the scientifically

guided development of an appropriate infrastructure for the collection, storage, care, processing and transfer of such donated data.

A5. Promote digital decision-support systems in clinical practice
Both the promotion of reciprocal knowledge transfer between research and clinical practice, and the approval of digital services to support decisions that can improve patient care should be accelerated. To this end, while safeguarding data sovereignty, it is necessary to grant legitimised actors the widest possible access both to data deriving from research and healthcare provision and the appropriate health-relevant big data applications.

A6. Promote international connectivity
With a view to international data exchange, efforts at standardisation should not be limited to national territories. Rather, far-reaching efforts must be made at all levels (policy, science and technology development) to align standards.

In order to promote the international competitiveness of German and European digital applications in the healthcare sector, including the high standards of quality- and data-protection this sector demands, and in order to counter problematic dependencies in this sector, investments in medical informatics should be far broader in scope and more swiftly implemented than previously planned. In particular, a targeted advancement of data management in public hospitals seems eminently sensible.

B. Ensure individual freedom and privacy

The willingness to place one's personal data at the disposal of third parties must be understood as part of one's informational freedom as a data provider. Data providers thus need to be equipped with the ability to handle their data in a sovereign manner and to consciously shape their private spheres. Furthermore, a framework must be created that guarantees appropriate scope of action.

B1. Safeguard data providers' sovereignty over their personal data
In light of big data's ability to recombine data and to detach it from specific purposes, the power of determination exercised by the data provider over their personal data must be safeguarded as comprehensively as possible.

B1.1. Open programmatic interfaces to data providers ("data agents")
Especially in situations where the scope of data usage cannot be precisely demarcated in advance, or when data collection and processing is continuous, appropriate software tools ("data agents") should be made available as a supplement to commonly employed consent models. These would administer the data feed according to the expectations of the data provider, thus enabling greater control, transparency and traceability. The corresponding programmatic interfaces should be standardised by self-regulatory or legislative means to facilitate the development of such data agents. The correct functioning of the interfaces and data agents should be supported by auditing or certification measures.

B1.2. Facilitate data providers' co-determination of data dissemination
When data is disseminated, the reversibility of data collection should be ensured: any system that collects personal data and accepts its input must – except for well-founded exceptions – be able to completely or partially delete this data. Here too, a model of data agents integrated as monitors into data pipelines should be deployed. Through suitable channels of communication (such as a corresponding app), the data provider should be asked to consent to the dissemination of their data and, depending on the case, be able to restrict or revoke it with relative ease.

B1.3. Clarify legal problems surrounding the supposed ownership of data
Data sovereignty should not be confused with the "ownership" of data. Insofar as the concept of ownership implies its essential legal elements – a permanent, fixed relationship and the absolute power of exclusion vis-à-vis third parties – it is poorly suited to the task of ensuring data sovereignty. On the other hand, because a certain (though flexible) personal sovereignty over data on the part of the individual must be recognized, it makes sense

to focus instead on the legal framework for the use of data. The German Ethics Council recommends establishing a comprehensive expert commission on this subject, informed not only by legal expertise but also by an interdisciplinary approach.

B2. Establish cascading consent models

In principle, a consent-based regulatory concept should continue to be applied in clinical practice and medical research (the opt-in model). Furthermore, cascading consent models should be employed whenever possible, to offer varied, dynamic ways of providing or delegating consent (e.g. to an independent fiduciary/institution or similar entity) – once, regularly, or for each individual decision. Provided that a basic attitude of respect towards the private sphere, alongside the safeguards and quality standards elaborated in this Opinion, are guaranteed, models that have proven effective, especially in the field of biobanks, should be transferred and adapted to other sectors.

B3. Ensure privacy-friendly default settings

Whether due to a lack of time or understanding, a subjectively perceived lack of alternatives, or done in good faith, data providers often simply accept the default settings of data-collecting and data-processing applications. Default settings should therefore be technically developed, under legal safeguards, to offer adequate protection of privacy from the outset (privacy by design/privacy by default). This applies in particular to the as yet relatively unregulated domain of private-sector offers such as health-related apps for mobile devices and their associated sensors and surveillance devices. In addition to the provisions of the GDPR regarding user-friendly settings, additional information should be provided to ensure that users actually understand the consequences of changes to basic settings.

B4. Explain and make transparent the use of algorithms

Beyond existing legal information requirements, the objectives, functions, and mechanisms of data collection and any algorithms used should be comprehensible to non-specialists. While taking into account the need to

protect intellectual property, this information should include the following in particular:

- what user data is input into what analyses, predictive models and decision-making or selection processes, and what attributes are expressly not collected and input in order, for example, to avoid discrimination,
- what inferences, conclusions, predictions, selections or decisions are derived from and made by means of algorithms working with this data,
- if and how profiles of data providers are created and what expected validity can be provided by such derived variables,
- in what form anonymised personal data feeds into (statistical) models and who has the rights to its use.

B5. Counter deception and manipulation

A distinction must be made between open, transparent ways of influencing others and more problematic forms of covert intervention that circumvent the cognitive control of those addressed and targeted. The manipulative acquisition and use of data, which deceives the data provider with regard to, for example, the nature and purpose of data collection, and/or exploits their inability to understand its implications, is legally and morally inadmissible. Especially in social networks, apps and online games, both governmental authorities and the operators themselves must work more vigorously to counteract such trends.

B6. Promote digital education

A prerequisite for data sovereignty is a basic understanding of the significance and value of big data as well as its associated risks. Given that children also use digital applications and generate data, the necessary user competence should already be conveyed in school. Beyond the purely technical aspects of conventional strategies for digitalising the classroom, imparting this competence should be regarded and designed as a task cutting across all subjects. This will counteract the informational self-endangerment currently endemic amongst children and adolescents, and raise early awareness of the relevant legal, social and ethical implications. Being able to convey the necessary user competence should as such be an indispensable

part of future teacher training. Additionally, institutions providing adult education should continuously maintain accessible offerings for all ages in this area, while companies and institutions should conduct regular internal training.

B7. Strengthen discourse and participation
An ongoing public debate on big data should be more thoroughly fostered. To this end, the state must work to ensure the provision of reliable information and to establish participative processes. These should ensure broad public participation and open exchange with experts and professionals.

C. Ensure justice and solidarity

C1. Create fair access to digital services
Certain groups of users are regularly excluded from the advantages of digitisation as a result of, for example, educational barriers. In order to counteract such factors, not only are special informational and educational provisions necessary; care must also be taken to ensure that digital services are not from the outset designed in such a way as to be exclusive. This could be the result of incomprehensible or unnecessarily complicated means of operation or unnecessarily technical language. Software and user interfaces should be designed to be as barrier-free as possible.

C2. Uncover and prevent discrimination and stigmatisation
Steps must be taken to ensure that the expanded body of information provided by big data, on the basis of which healthcare-related allocation decisions can be made, is not abused such that certain persons or groups of people are subject to discrimination or stigmatisation. When applying the insights yielded by big data analysis, there exists an acute danger that the underlying data, the selected parameters of the analysis and/or the algorithms employed may produce results that entail systematic and insidious forms of discrimination against people or groups of people. For this reason, it is not only necessary to establish in advance the inadmissibility of certain, corresponding selection criteria – unless they have explicit and appropriate

purpose – but also to develop procedures by which possible violations can be identified and sanctioned. Even if subsidiary self-regulation by sector or by institutions themselves is effective in this regard, it must be supplemented by governmentally enforced, sanctioned and judicable safeguards.

C3. Allow objections to automated decisions

When dealing with decisions that are made based on algorithms, structural forms of objection and redress are necessary. Especially in the domain of private insurance, the policyholder's right to a clear, comprehensible and individual justification for a rejected claim for compensation must be guaranteed, as well as free and low-threshold access to internal and external appeal and arbitration bodies.

C4. Protect vulnerable individuals and groups

Special attention must be paid to individuals and groups who, because of individual or social circumstances, are (at least temporarily) more liable to be directly or indirectly, structurally or intentionally denied the benefits of, or made to disproportionately bear the costs of, digitisation in the healthcare sector. This applies especially to children and young people as well as to the elderly and people with disabilities. Not only must these individuals be supported in terms of developing their ability to use digital services responsibly, they must also, because of their specific vulnerability, be given special protection in the process of data collection and use. In this respect, data sovereignty also takes account of the individually and situationally varying capacity for responsibility on the part of those affected by big data.

C4.1. Strictly comply with the consent requirements for children and adolescents

The provisions of the GDPR regarding the consent of minors in relation to information society services should be swiftly and thoroughly implemented. Decisions regarding the option of lowering the minimum age of consent (allowed for by the GDPR) should not be made without the involvement of those concerned (i.e. children and adolescents).

C4.2. Develop data-protection mechanisms for others with limited capacity to consent

Special data-protection mechanisms should be developed to regulate the collection of data from other persons who have a limited capacity to provide consent, while not inhibiting the potential to conduct big-data-based research with and for the benefit of such individuals. Participating research institutions should ensure that information sufficient for informed decision-making is provided both to affected individuals according to their cognitive faculty, and caregivers, in line with the principle of decision-making assistance.

C4.3. Restrictively regulate the use of chatbots

The use of chatbots to collect data from or pertaining to persons with limited cognitive faculty entails an especially high potential for manipulation, and should therefore be regulated in a particularly restrictive way.

C5. Safeguard care-oriented medicine

Personal attention to and care of the patient in medical practice should be enhanced, not compromised, by the use of big data applications. Time and money saved by relieving care personnel of routine work or providing faster and more accurate diagnostics through digital algorithms should translate into an increased personal attention paid to patients.

C6. Ensure effective liability of companies working with data in the healthcare sector

Given the risks associated with big data, it seems appropriate to develop specially adapted liability models. Here, it must first be established whether and to what extent the new regulations contained in German data protection law (which thus far do not exhaust the possibilities of the GDPR) are sufficient. The GDPR provides for the introduction of strict liability to provide the individual with effective protection from damages. Given the uncertainties of liability and the rules of evidence, such a form of strict liability, tailored to the specific risks of big data, should be considered. Independent of an application's authorisation, this liability should be excluded

only if the damage is inevitable. The amount of any potential limitation of liability should be set high enough to exert a noticeable effect on large companies.

D. Promote responsibility and trust

D1. Guarantee protection and quality standards

D.1.1. Establish the best possible standards of protection against the unauthorised identification of individuals from anonymised, pseudonymised or aggregated datasets

Given the inadequate protection offered by established anonymisation and pseudonymisation techniques, adequate complementary safeguards should be established to mitigate the risk of re-identification:

- Where identifiers (e.g. e-mail, login, device ID, cookie ID) allow for relatively direct inferences regarding affected persons, these must be replaced by anonymised keys designed to expire as quickly as possible.
- Whenever an anonymous user directly or indirectly reveals their identity, either unexpectedly or accidentally, (e.g. the accidental disclosure of one's name, e-mail, telephone numbers, credit card number, ID number, etc.) the onus is on the data collector to ensure that this identification is reversed through data deletion.
- Wherever a dataset, through a combination of attributes and data, renders a user identifiable with a high degree of probability, this data should be subject to the same data protection regulations as explicit identifiers.
- If connections between datasets entail a certain reduction in the level of protection, these datasets must be isolated or linked only briefly (i.e. without permanent storage in a database) and for well-defined purposes.

D.1.2. Compensate for anonymisation gaps by controlling access to data

Given the persistent risk of re-identification, controlling access to data is of particular importance. Especially in clinical practice and basic medical

research, data access must be appropriately restricted to authorised parties. This should be accomplished by storing data in secure, technically isolated and independent repositories, and establishing controlled means of access, including robust verification and authentication systems.

D.1.3. Ensure and certify the implementation of protection requirements
Data sovereignty requires the coexistence of technical and regulatory standards. In addition to existing privacy-by-design guidelines, anyone processing and using data must place strive to ensure that privacy-related considerations are a top priority for any project making use of big data, beginning with the planning and development phase. It should also be incumbent on the institutions concerned (in research, in medical practice or in the commercial field) to demonstrate compliance with regulations designed to secure data sovereignty in their respective fields of responsibility. In addition to their existing experience, internal data protection officers should further develop their areas of responsibility and authority in this direction (corporate data governance).

D1.4. Establish reporting requirements for mishaps and misconduct
Care must be taken to ensure that any mishaps and misconduct in the handling of data do not remain hidden, but are instead understood in terms of their relevance to the entire system and productively learned from. This would entail a duty to inform potentially injured users and, to the extent that these cannot be identified, the public, as well as to report irregularities to the supervisory authorities/bodies.

D2. Improve control mechanisms

D2.1. Strengthen the role of data protection officers
Ensuring data sovereignty requires an array of internal (private) and external (state) supervisory authorities, whose responsibilities should be better demarcated and whose capacities and expertise should be broadened where necessary. In particular, it is both eminently sensible and necessary to reorient the roles of existing data protection officers – in both the public

and private sectors – towards the goal of data sovereignty and, if necessary, to extend their roles. They should complement the work of local supervisory bodies such as research ethics committees, and should moderate and arbitrate, on the basis of transparent decision-making criteria, in conflict situations. Insofar as the existing control structures prove inadequate in addressing the specific problems raised by big data, for example in multi-regional and international joint projects, greater centralisation should be considered.

D2.2. Establish data auditors
In light of the fact that data quality is important to society as a whole, especially in medical research and clinical practice, a corresponding review and control structure should be established. This would not necessarily have to involve an exclusively public authority; it could also be designed as private regulation, analogous, for example, to financial auditing and accounting in corporate law.

D2.3. Introduce data guardianship models
In order to foster trust and prevent abuse, those using data should lay the technical and organisational groundwork for ensuring that data stocks are not necessarily given directly to them, but that models of guardianship (such as charitable trusts) can be interposed. Not only can this mitigate power imbalances; it can also counteract conflicts of interest. At least in the area of medical research and clinical practice, policymaking steps should be taken to ensure that such models are effective, particularly with regard to data users operating in an international context (e.g. Google, Apple, Facebook, Amazon and Microsoft).

D3. Develop codes of conduct for research, clinics and industry
Using existing codes of conduct as a model, consistent and sustained effort should be made to establish comprehensive internal standards of conduct in all areas sensitive to data-protection issues. This must involve not only taking account of the applicable regulatory requirements and strengthening them where necessary, but also – at least within industry or in connection

with specific fields of application – to strive for coordination and alignment across national borders.

D4. Support and expand quality seals for service providers and applications

Because ensuring compliance with the principle of data sovereignty is in the best interests of data users, a system of market-based classifications (quality seals), some of which already exist, should be supported and expanded. In this way, efforts beyond achieving minimum standards and meeting compulsory legal requirements can help profile a company vis-à-vis the competition. Insofar as these self-regulatory mechanisms prove inadequate, co-regulatory measures – for example in the form of official certifications – must be introduced. Structures of state control, including liability provisions, must also be strengthened.

D5. Strengthen competence in the responsible handling of data among everyone involved professionally with big data

In fields where big data's role is rapidly expanding, it is necessary to promote awareness of the ethical challenges and new responsibilities arising from the use of health-relevant data. To effect such a cultural change requires an improved understanding of research and information ethics among all parties involved, as well as the ability to scientifically and critically reflect upon one's own actions. Imparting these competencies should become a mandatory element in professional training, as well and higher and further education, touching on all relevant subjects and fields. To do justice to the complexity and significance of this issue, for example, companies and institutions could expand their efforts to establish internal data science departments.

>> DISSENTING VOTE

In her dissenting vote, Christiane Fischer calls for the renunciation of the use of big data for research purposes or other applications if comprehensive data protection, the implementation of effective anonymisation and pseudo-anonymisation standards, and the right to forget cannot be guaranteed.

Members of the German Ethics Council

Prof. Dr. theol. Peter Dabrock (Chair)
Prof. Dr. med. Katrin Amunts (Vice-Chair)
Prof. Dr. phil. Dr. h. c. Dipl.-Psych. Andreas Kruse (Vice-Chair)
Prof. Dr. med. Claudia Wiesemann (Vice-Chair)

Constanze Angerer
Prof. Dr. iur. Steffen Augsberg
Prof. Dr. theol. Franz-Josef Bormann
Prof. Dr. med. Alena M. Buyx
Prof. em. Dr. iur. Dr. h. c. Dagmar Coester-Waltjen
Dr. med. Christiane Fischer
Prof. em. Dr. phil. habil. Dr. phil. h. c. lic. phil. Carl Friedrich Gethmann
Prof. Dr. rer. nat. Dr. phil. Sigrid Graumann
Bishop Prof. Dr. theol. Martin Hein
Prof. Dr. med. Wolfram Henn
Prof. Dr. iur. Wolfram Höfling
Prof. Dr. (TR) Dr. phil. et med. habil. Ilhan Ilkilic
Prof. Dr. rer. nat. Ursula Klingmüller
Stephan Kruip
Prof. Dr. phil. Adelheid Kuhlmey
Prof. Dr. med. Leo Latasch
Prof. Dr. iur. Dr. h. c. Volker Lipp
Prof. Dr. theol. Andreas Lob-Hüdepohl
Prof. em. Dr. iur. Reinhard Merkel
Prof. Dr. phil. Gabriele Meyer
Prof. Dr. med. Elisabeth Steinhagen-Thiessen
Dr. phil. Petra Thorn

Office

Dr. rer. nat. Joachim Vetter (Head of Office)
Dr. theol. Katrin Bentele
Carola Böhm
Malica Christ
Ulrike Florian
Dr. phil. Thorsten Galert
Steffen Hering
Christian Hinke
Petra Hohmann
Torsten Kulick
Dr. Nora Schultz