Deutscher Ethikrat

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

OPINION

Deutscher Ethikrat

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

OPINION

30 November 2017

Published by the German Ethics Council

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

© 2026 Deutscher Ethikrat, Berlin
Title of the original German edition: Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung

All rights reserved.
Permission to reprint is granted upon request.
English translation: Forest Holmes
Layout: Torsten Kulick

ACKNOWLEDGEMENTS

The German Ethics Council sincerely thanks for their commitment all those who have contributed to this Opinion, either in writing or orally, in the form of inspiration or information.

The project was launched with the annual meeting "The Quantified Human – Big Data and Health" on 21 May 2015. With contributions from the invited speakers Elisabeth André, Anke Domscheit-Berg, Arno Elmer, Nils Hoppe, Christof von Kalle, Peter Langkafel, Klaus Mainzer, Wolfgang Marquardt, Günther Oettinger, Thomas Petri, Frank Rieger, Florian Schumacher, Stefan Selke and Henry Völzke, as well as numerous discussion contributions from its more than 500 participants, this conference provided an excellent basis for the project going forward.

Together with Christoph Kucklick, Stefan Selke met a second time with the German Ethics Council for a public plenary meeting on 23 March 2016. Jan Philipp Albrecht, Urs-Vito Albrecht, Luciano Floridi, Thomas Hofmann, Sascha Lobo, Indra Spiecker, aka Döhmann, and Bart de Witte also brought their expertise to internal hearings.

From February to March 2017, numerous interested parties and organisations participated in a public consultation on big data and health, the results of which were incorporated into further deliberations of the German Ethics Council and will be published separately to accompany this Opinion.[1]

Up to the end of their term of office (2016), Council members Wolf-Michael Catenhusen, Thomas Heinemann, Anton Losinger, Ulrike Riedel, Eberhard Schockenhoff, Jochen Taupitz, Christiane Woopen and Silja Vöneky also participated in the working group dealing with the preparation of this

1 All publicly available accompanying information and documentation belonging to the German Ethics Council on this subject can be accessed at http://www.ethikrat.org/themen/forschung-und-technik/big-data.

Opinion. Sascha Lobo (2015) and Thomas Hofmann (from 2016) provided valuable support as co-opted working group members.

Special thanks are due to the students who, during the course of the project, made decisive contributions to the completion of this Opinion with their dedicated work during internships at the German Ethics Council.

CONTENTS

EXECUTIVE SUMMARY 9

1 INTRODUCTION 44

2 BASICS: BIG DATA AND HEALTH 49

2.1 The characteristics of big data 50

2.2 The collection and handling of large volumes of data 52

2.3 Data analysis and data science 58

2.3.1 Statistical modelling and validating correlations and mechanisms of action 59

2.3.2 Machine learning and machine perception 66

2.3.3 Stratification and individualisation 73

2.4 The personal and health-related nature of data 74

2.4.1 Data of a personal nature 75

2.4.2 Data of a health-related nature 77

2.4.3 Decontextualisation and recontextualisation 79

2.5 Actors and contexts of action 82

2.5.1 Big data in biomedical research 82

2.5.2 Big data in healthcare 93

2.5.3 The use of health-related data by insurers and employers 97

2.5.4 Commercial use of health-related data by global IT and internet companies 103

2.5.5 Collection of health-related data by individuals 106

2.6 Interim conclusion 110

3 LEGAL GUIDELINES FOR BIG DATA 112

3.1 Control directives grounded in basic rights 113

3.2 Provisions in non-constitutional law 116

3.2.1 Big data as a challenge for applicable data protection law 117

3.2.2 Health data protection law 130

3.2.3 Interim conclusion 133

3.2.4 Medical device law 134

3.2.5 Big data services in the context of (statutory) health insurance 138

3.3 Regulatory options 141

3.3.1 Further development of existing laws 144

3.3.2 The regulatory function of private law 148

3.3.3 The possibilities of cross-border regulation 151
3.3.4 The supplementary function of non-government control instruments 154
3.4 Conclusion: Statics and dynamics of the legal framework 157
4 ON THE ETHICS OF BIG DATA AND HEALTH 158
4.1 Freedom: authorship and self-determination 160
4.1.1 Authorship of one's actions 160
4.1.2 Self-determination and consent 163
4.1.3 External parameters for the realisation of freedom 170
4.2 Privacy and intimacy 172
4.3 Sovereignty and power 182
4.4 Non-maleficence and beneficence 195
4.5 Justice 201
4.6 Solidarity 207
4.6.1 Solidarity in statutory and private health insurance 211
4.6.2 New practices of solidarity 217
4.7 Responsibility 219
4.7.1 Individual responsibility as regards the sharing of health-related data in various roles and contexts 221
4.7.2 The responsibility of institutional actors 222
4.7.3 The responsibility of state bodies 227
4.7.4 Conclusion: multi-actor responsibility 229
5 DATA SOVEREIGNTY AS THE SHAPING OF INFORMATIONAL FREEDOM 230
5.1 Data sovereignty as a guiding concept 231
5.2 Data sovereignty in the healthcare sector 233
5.3 Outline of a governance concept oriented towards data sovereignty 235
6 RECOMMENDATIONS 240
DISSENTING VOTE 258
REFERENCES 263
CITED DECISIONS 279
ABBREVIATIONS 280

EXECUTIVE SUMMARY

Fundamentals: Big data and health

1) Big data is one of the key terms in present debates concerning 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. In light of the volume and variety of the data being processed, as well as the velocity with which this data is captured, analysed and linked, this requires innovative and continuously refined approaches and technologies.

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-related data by individuals, for example by means of smartphone apps and sensors worn on the body, is constantly increasing. The linking 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 linked, often across national 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 any 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 (causal effects) or modes of action. These can only be elucidated through additional arguments and assumptions or by means of additional data, for example 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 allows 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 decontextualisation and recontextualisation of data that is 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 deanonymised, or individual persons be reidentified.

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 its 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-related 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-related 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 easily 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 their 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. Drawing upon large volumes of data, healthcare professionals 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 within 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 programmes 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 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-related 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 the 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-related 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-related 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 Bundesverfassungsgericht (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 personality 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 have not rendered it sufficiently adapted to the phenomenon of big data. This is true 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 this data is collected.

44) Moreover, beyond the point of consent, existing data protection law offers 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 linking of diverse data, big data increases the chances of reidentification and undermines the effectiveness of anonymisation and pseudonymisation requirements. To what extent and at what point is the danger of the reidentification 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 the 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 view 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 that predates 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 General Data Protection Regulation 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 will 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-related 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 conducting 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, recent 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 in which consent is repeatedly sought with regard to individual elements of data use are complemented by further options, such as the possibility to delegate authorisation. Participants can furthermore decide 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 sense, 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 privacy 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 regard 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 heteronomy 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 is 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 together 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 they, 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 possess 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 their 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 harm, 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 healthcare sector.

70) Knowledge and insight are of major importance in the self-constitution of the individual and their 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 process of data use 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-related data allows for the generation of more precise risk profiles. The concern thus arises that the basic assumption of solidarity underlying the statutory health insurance system as well as the fair structuring of contracts in the private health insurance industry – namely, the assumption 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 publicly 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, however, 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 to 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 one's 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 concerning 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 era of big data, 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 applies, 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 decontextualisation 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 set up delegation systems at the programmatic interfaces in data networks. 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 tamper-proof recording 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, quality seals, 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 above-mentioned 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 this 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 to have 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 growing role of 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 of 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 emphasise 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 invokes 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 the 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 there.

Recommendations

104) 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.

105) The governance concept presented here contains specific recommendations for action in four areas. These aim, firstly, to realise the potentials of big data; secondly, to preserve individual freedom and privacy; thirdly, to ensure justice and solidarity; and fourthly, to foster responsibility and confidence. The measures recommended here should be financed and implemented as soon as possible.

106) Recommendations for developing the potentials of big data in the healthcare sector (area A) touch on the following points:
   - Creating the infrastructural prerequisites (A1)
   - Facilitating data exchange and integration (A2)

- Developing and making available standardised procedures for the interoperability of data (A2.1)
- Further developing cooperative data management in research (A2.2)

» Promoting and preserving data and research quality (A3)
- Maintaining epistemological standards, particularly those of evidence-based medicine (A3.1)
- Introducing uniform data and documentation standards (A3.2)
- Establishing quality seals for data (A3.3)

» Adapting the legal framework for data use in research (A4)
- Further developing the secondary use of research data (A4.1)
- Creating the legal conditions that allow individuals to consent to the comprehensive use of their data for medical research ("data donation") (A4.2)

» Promoting digital systems to assist decision-making in clinical practice (A5)

» Promoting international competitiveness and connectivity (A6).

107) Recommendations for securing individual freedom and privacy (area B) encompass the following aspects:

» Preserving users' sovereignty over their data (B1)
- Opening programmatic interfaces for data providers ("data agents") (B1.1)
- Facilitating choice throughout the data sharing process (B1.2)
- Clarifying the legal problems surrounding a supposed ownership of data (B1.3)

» Establishing cascading models for granting consent (B2)

» Guaranteeing default settings that favour privacy (B3)

- Explaining and rendering transparent the use of algorithms (B4)
- Countering fraud and manipulation (B5)
- Promoting digital education (B6)
- Strengthening public discourse and participation (B7).

108) In order to preserve justice and solidarity under big-data conditions (area C), the German Ethics Council recommends:
- Creating fair access to digital products (C1)
- Uncovering and preventing discrimination and stigmatisation (C2)
- Enabling objections to automated decisions (C3)
- Protecting vulnerable groups and individuals (C4)
  - Strictly observing requirements for consent from children and youth (C4.1)
  - Developing protective mechanisms for data collection from other persons with limited capacity for consent (C4.2)
  - Restrictively regulating the use of chatbots (C4.3)
- Ensuring that medicine remains oriented towards care (C5)
- Guaranteeing effective liability on the part of businesses working with data in the healthcare sector (C6).

109) The following recommendations should foster responsibility and confidence in the health-related application of big data (area D):
- Guaranteeing standards of protection and quality (D1)
  - Establishing the best possible standards of protection against unauthorised identification of individuals from anonymised, pseudonymised, or aggregated datasets (D1.1)

- Compensating for anonymisation deficits through monitored access to data (D1.2)
- Ensuring and verifying the implementation of data protection provisions (D1.3)
- Establishing an obligation to provide information in cases of malfunction or abuse (D1.4)

» Improving control mechanisms (D2)
- Strengthening data protection officers (D2.1)
- Establishing data examiners (D2.2)
- Introducing data trustee models (D2.3)

» Developing codices for research, clinics and industry (D3)

» Supporting and expanding quality seals for applications and their suppliers (D4)

» Strengthening competence in the responsible handling of data amongst everyone working with big data professionally (D5).

1 INTRODUCTION

Big data is among the key terms in present debates concerning the societal changes being driven by technological change. Although the various technologies grouped as big data often do not share a single, established definition, the term nonetheless has exerted a powerful effect on public discourse. Behind the phrase big data lies one of the central mechanisms in the world of digital data: the collection, analysis and linking of ever-larger quantities of information by means of a sophisticated processing infrastructure. The total amount of data in global circulation is constantly and rapidly increasing; through recontextualisation and recombination, new insights are constantly being gleaned both from existing and from newly generated data. The processes associated with big data thus far exceed the scope of conventional data collection, which is restricted to specific purposes. They challenge the approach we have thus far taken towards data and confront us with new questions regarding a number of social practices – such as the drawing up of insurance policies – and appropriate legal standards, specifically data and privacy protections. The discussion surrounding big data goes to the core of our self-perception both as individuals and as a society being shaped by digitisation and increasing interconnectedness.

The transformative potential big data holds for both individuals and society can be seen with particular force and clarity in the healthcare sector. Growing numbers of researchers, firms and doctors already work with massive quantities of data. However, the collection of health data has long since ceased to be restricted to doctors' practices and medical studies; it is now generated by citizens themselves, by means, for example, of sensors and apps on their smartphones or devices worn on their bodies (smartwatches, fitness trackers). At the same time, developments in data science and infrastructure are opening up new possibilities for quickly and effectively

analysing, exchanging and linking diverse data both across datasets and with other information, the relevance of which to health may only become apparent when viewed in a shared light. This could include, for example, information regarding one's shopping behaviour, online search queries, location data or the analysis of text and recorded material. Such analyses not only facilitate a deeper glimpse into one's state of health, personality and lifestyle; they even allow predictions to be made. Overall, an ever-expanding number of data sources are being fed into big-data-based fields of analysis.

This linking of various kinds of data promises to open up new avenues in scientific research and medical treatment strategies. The combined analysis of, for example, classical medical data, research data, public healthcare data, movement data, fitness data, data from social networks and insurance data has allowed for wholly novel insights and approaches to confronting health risks and diseases. That such visions of better, faster and more precise diagnosis, prevention and treatment at all appear possible, however, also has to do with the blurring of boundaries in our thinking about health questions, by which everything that we do and experience is subject to analysis not only in connection with our health, but also in determining our status as sick or healthy. This applies to medical diagnoses and options for therapy and prevention, as well as to the calculation of insurance premiums or occupational risks.

Health data are conventionally regarded as especially sensitive information and are correspondingly protected. There are at least two reasons for this. Firstly, when health conditions regarded as unfavourable come to light in a given social context (in work, contracts, relationships), this can entail disadvantages for those affected – including discrimination and stigmatisation – and can result in social and, where relevant, contractual and financial exclusion. Secondly, beyond such practical considerations, one's physical constitution is regarded in our society as something profoundly intimate. Apart from in certain exceptional, individually justified instances, it

is something for which one can never be called to account by others. Whether any given data can be considered sensitive or health-related, however, is often impossible to determine when it is gathered in connection with big-data applications, with their novel capacities for combination and analysis. Rather, this depends increasingly on the context in which the information is employed.

The massive growth in influence in this sector of large IT and internet firms, which are based outside of Europe and remain, even within Europe, difficult to regulate, only intensifies the challenges faced by individuals and society. Not only do central principles such as privacy and informational self-determination appear ever more difficult to defend against outside attack, they are also undermined by the behaviour of affected individuals themselves, who willingly accept violations of the most intimate personal domain in return for the services offered by numerous apps, programs and internet-based technologies. This points to another area of difficulty: individualisation, which holds the potential of being enormously beneficial in the areas of prevention and treatment, could, in the insurance business and the wider social sphere, summon conscious or unconscious dynamics erosive to social solidarity. A reduction in one person's insurance premium could ultimately translate into a financial disadvantage for another. Anyone unwilling in the future to follow popular behavioural guidelines could run the risk of only obtaining insurance cover under disadvantageous conditions, if at all.

Big data in the healthcare sector thus offers promising opportunities and prospects for research and the prevention, diagnosis and treatment of disease. On the other hand, however, it presents us with serious challenges and risks impacting various aspects of society. In light of the major changes these opportunities and risks could entail, both for the individual as well as for society as a whole, numerous questions are raised by the deployment of big data in connection with health. How, for example, will the ever more fine-grained and often barely

noticeable collection of health-related information affect our freedom, self-perception, and self-determination? Will that which appears under big-data conditions to be a form of self-determination ultimately reveal itself as a kind of self-disenfranchisement, or is it simply a shift in one's conception of oneself that is at stake, such as often occurs over the course of a lifetime? How can individuals, researchers and companies deal responsibly with big data, and how can the quality and reliability of complex data analyses be ensured? Which challenges are posed to solidarity and justice, when considering, for example, health insurance, but also research, which, when employing big-data analytics, requires massive quantities of data? How can confidence in research be maintained if even anonymised participant data could, in principle, always be decoded? What regulatory mechanisms and incentives could appropriately manage the risks and opportunities presented by big data in the healthcare sector? And where, from the perspective of the individual or the regulating state, should lines be drawn to limit the collection, linking and use of data? Should we set any limits at all, or should we loosen existing restraints, such as data protection standards, in order to better take advantage of the desired benefits of big data?

Faced with such questions and challenges, the task at hand is to at least focus our attention on the extent to which big data influences understandings of health, self-determination and social interactions. The goal must be to shape this technological development so that it does not simply generate profits for the companies involved – which in itself is wholly legitimate – but also safeguards and promotes the real ability of society's members to conduct and determine their lives as free subjects. Against this backdrop, the Opinion presented here seeks to stimulate awareness of the urgent issue of health and big data, while setting ethical and legal standards in this field.

It begins by outlining the scientific and technical as well as the economic basis of big data in the health sector (chapter 2). It then moves on to describe the relevant legal framework

and to identify regulatory gaps (chapter 3). Fundamental ethical questions are addressed next (chapter 4). The answers to the latter culminate in the call for a governance concept that is tailored to the novel conditions presented by big data and aims to achieve the central goal of data sovereignty (chapter 5). The meaning of this term in connection with big data varies. Here, data sovereignty is understood as the responsible shaping of informational freedom, in a manner appropriate to the risks and opportunities presented by big data. This concept thus takes up fundamental legal and ethical considerations and develops them further with respect to the specific domain in question. The task of realising data sovereignty is a complex one, whereby a multitude of actors and diverse interests must be taken into account. The foundations for this task now being laid – technically, economically, legally, politically and culturally – will decide the long-term meaning of health and self-determination in our society. In this light, the German Ethics Council concludes this Opinion by formulating its recommendations for dealing with the recognised ethical, legal and social challenges posed by big data in the health sector (chapter 6).

2 BASICS: BIG DATA AND HEALTH

Since the early modern period at the latest, the systematic collection and analysis of data has represented a decisive factor in civilisational development. Systematic experiments, measurements and observations form the basis of the empirical sciences. The history of science has been characterised by the invention of new instruments and improved procedures that can be used to measure our world. Without these inventions, neither the successful formation of scientific theory, nor lasting scientific, technological and economic progress would be conceivable. Data collection and information gathering also play an important role in addressing problems of social self-organisation, such as those that arise in the healthcare system.

The scientific measurement of the world is a project that also encompasses humans and their living environment. In the medical context, this ranges from determining physical parameters such as pulse, temperature, respiration, or blood pressure, to modern laboratory diagnostics, imaging techniques and genome analyses. Likewise, personality traits, emotional states and cognitive abilities or motivations are the subject of empirical research aimed at explaining, rendering more predictable or even influencing human behaviour. Psychometrics, for example, aims to measure psychological phenomena while attempting to link (experimentally) measurable data to presumably underlying factors. Epidemiology and the social sciences extend the empirical method to groups or social collectives, often relying on available data taken from registers, archives and clinical databases, but also collecting new data by observing or questioning subjects.

The computer first appeared as an instrument for performing mathematical computations and, later, electronic data processing (EDP). In the sciences, the use of computers enabled an increase in the volume of manageable data, as well as numerous qualitative improvements, such as the use of more

complex algorithms (models) in computationally demanding computer simulations. Added to this was the revolution in business processes driven by the introduction of the computer. Initially, EDP was often limited to the manual collection of data and its storage, maintenance and aggregation. The characteristics of classical EDP were – and still are – work processes and a business logic that interweaves human activity with limited automated steps, usually by means of special software (e.g. in accounting, inventories, customer management, etc.). This also applies to the domain of healthcare, for example when information technology (IT) is employed in hospitals, in medical practice or in the operations of insurance companies. In this context, data is almost always structured, i.e. rendered interpretable within a database scheme that sets measurement protocols, data formats and data semantics. Rationalisation, standardisation and quality improvement are the main objectives of this technological development.

Since the 1990s, decision-making processes have been ever more thoroughly underpinned by systematic analytical methods and central data warehouses optimised for this purpose. With their help, various types of data are extracted from their original sources and processes, merged and condensed on a large scale, and finally made available to data analysis, which employs increasingly complex algorithms.[2]

2.1 The characteristics of big data

The evolution towards big data has been accompanied by a transformation of all phases of data processing, characterised by increasing automation, networking and pervasion. It is difficult to arrive at a precise definition of big data; the blurred

2 Algorithms are unambiguous sets of mathematical instructions and procedures that consist of defined steps and lead to a specific goal; they represent the basis of computer programs.

boundaries of the term are sometimes even regarded as a characteristic aspect of the phenomenon.[3] Nonetheless, big data refers not only to the sheer quantity of data processed, but also to the qualitative leaps in both the requirements for and potential of data analysis. The German Ethics Council bases its recommendations on the following working definition:

> Big data is the handling of large volumes of data in order to identify patterns and thus gain new insights. To this end, and because of the massive volume and variety of data, as well as the velocity at which this data is captured, analysed and recombined, big data employs continuously evolving information technologies and techniques.

This definition contains three key terms that appear in many definitions of big data[4] and which are known as "the three Vs": volume, variety and velocity. Further "Vs", namely validity or veracity (assurance of data quality and credibility) and value (in the commercial sense) are also often mentioned as important target metrics of big data.[5] Of special relevance to the – particularly ethically sensitive – domain of healthcare addressed here is the comprehensive decontextualisation and recontextualisation of data, which is collected, analysed and recombined for different purposes, enabled by big-data technologies (see section 2.4). Because of the steadily expanding volumes of data and possibilities of recombination, as well as the dynamic development of technical capacities, the concept of big data is undergoing constant change, for example in the

3 Cf. Mayer-Schönberger/Cukier 2013, 6.

4 Cf. e.g. the definition in the Gartner Information Technology Glossary: "Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation." (http://www.gartner.com/it-glossary/Big-Data [2017-10-17]). See also Laney 2001.

5 See e.g. Amma 2016, 54. Ultimately several other Vs can be identified in the discussion surrounding big data. Cf. here, e.g., Cartledge 2017.

field of machine learning. However, the elements outlined above are likely to remain relevant for a definition of big data.

2.2 The collection and handling of large volumes of data

As a result of technological progress, ever larger volumes of data from an ever larger variety of sources are being made available with ever increasing velocity. This is mostly a result of increasing automation. The proverbial EDP expert sitting in front of an input mask and moving from input field to input field has vanished in the world of big data. Many scientific and technological advancements in fields ranging from basic physical theory to materials science to nanotechnology, as well as a wide variety of new manufacturing techniques, contribute to the development of novel instruments, measuring equipment and procedures. These have at least one thing in common, namely to push data acquisition forward to previously unknown extents. Rapidly captured datasets several petabytes[6] in size are no longer a rarity, and the trend of data growth appears relentless.[7] The volume and speed of fully automated data collection have increased by many orders of magnitude in only a few years; they now seem to lie well beyond the limits of human imagination. Estimates place the total volume of data worldwide in 2016 at around 16,1 zettabytes (equivalent to 16.1 trillion gigabytes); with an annual growth rate of 30 percent, this volume is projected to reach about 163 zettabytes in 2025.[8] Up to 80 percent of this data consists of so-called unstructured data, such as videos, sound recordings, e-mails, research data and posts on social media.[9]

6 A petabyte corresponds to 10^{15} bytes, or one million gigabytes.
7 See e.g. Cisco 2017.
8 Cf. Reinsel/Gantz/Rydning 2017, 3.
9 Cf. Stone 2014, 2.

As a consequence of the enormous increase in the amounts of stored data, human operators often can no longer directly control its quality. This applies, for example, to errors in measurement, outliers and other irregular data contained in large datasets, such as the databases kept by hospitals, which demand new methods of analysis and automated quality assurance procedures. On the other hand, it is sometimes necessary to integrate humans into the process of data collection or analysis; this then involves the use of crowdsourcing, by which well-defined questions, tests or simple analytical tasks are passed on to human "experts". This reversal of the relationship between man and machine, in which humans serve to assist a computer program, is also called human computation.

In addition to the increasing use of powerful medical and research equipment, these large volumes of data also result from the fact that human life is being shaped to an ever greater degree by digital technologies. The rapid dissemination and networking of devices that, either as their primary function or incidentally, serve or can be used to collect data, has allowed for the constant tapping of new sources of data. Alongside data collection by means of the full range of computers (PCs, notebooks, game consoles, tablets, smartphones), this also includes the use of portable devices (fitness trackers, smartwatches) and networked home appliances to record bodily functions and personal habits; the monitoring of public space by, for example, cameras and various sensors that orchestrate the operation of machines (from cars to industrial plants); the measurement of environmental variables (through remote sensing); the movement of goods, and more.[10] The smartphone alone can rightly be regarded as one of the most versatile and

10 The movement of goods can be tracked using radio-frequency identification (RFID) systems that enable the identification and location of equipment by means of electromagnetic waves. Items equipped with RFID transmitters (transponders) can remotely exchange data with RFID receivers (readers) located within a reception area. An especially simple application of this technology is theft prevention in department stores.

widely used measuring instruments in human history. Every day smartphones capture movement data (GPS, WPS), web and app usage, search queries (Google, Siri), communication (calls, SMS, Messenger, e-mail, social networks), transactions (purchases, bookings) and much more billions of times over. This is all data with a clear connection to individual persons, and it allows conclusions to be drawn regarding the persons from whom it originates. From the users' perspective, it is often the case that much of this data is regarded as incidental or a mere byproduct of their daily activities and its storage or transfer is neither intended nor even considered.

Once data is captured, it is exchanged and linked by means of data networks and networked software systems, often in real time, in a scalable manner, and on a transnational or even global scale. Widely employed standards exist for the exchange of data and messages over the internet, mostly through application programming interfaces (APIs) that build on the Hypertext Transfer Protocol (HTTP). These standards facilitate the programmatic exchange and linkage of data between different computer systems, even with the involvement of multiple parties. An essential aspect of programmatic data exchange is processing consistency: once the interfaces are implemented and the systems mutually authorised, data connections are established that allow data to flow as if through a complex system of pipes within a very short timeframe and without the need for further human intervention.

Programmatic interfaces and the blockchain: technically supported transparency and options for control in data processing

The automated exchange of data over the infrastructure described here is not fundamentally inconsistent with the safeguarding of the individual's decision-making power over data. The same technical instruments, particularly the use of programmatic interfaces and the automatic logging of the origin of data, present a largely unused potential: appropriate software tools can support data providers in continuously managing the data they make available in accordance with their own wishes (e.g. with respect to storage, erasure, noise reduction, anonymisation). These

tools can function as proxies or agents on behalf of the data provider, in that they implement individual or stakeholder-driven rules and guidelines for the handling of data. As data processors and users today often rely on such models for data exchange, it would be possible to achieve greater control, transparency and traceability with limited technical effort. Blockchain technology, for example, which has received attention mostly in connection with the digital currency Bitcoin[11], allows the decentralised exchange of information between parties while providing a transparent documentation of all transaction steps impervious to manipulation. In a blockchain, transactions are added as data blocks, using cryptographic methods, to a constantly expandable list that functions as a decentralised accounting ledger. In addition to a time stamp and the transaction data, each block contains a cryptographically secure key to the preceding block. Since each block builds on earlier blocks, any previously recorded information or transaction cannot be subsequently manipulated. For the healthcare sector, blockchain technology would thus allow an especially secure documentation of steps in the processing and transfer of data as well as of relevant metadata, for example regarding data origin and quality or consent to certain types of data use. In the United Kingdom, for example, the Alphabet subsidiary DeepMind is currently planning to employ blockchain-based methods to guarantee the integrity of (meta) data, with the aim of ensuring greater transparency in data use.[12]

The efficient capture, storage, and processing of data requires a powerful computing infrastructure, most of which is provided in data centres. Vast amounts of data running into the multi-petabytes place high demands on the capacity of memory, networks, interfaces and the underlying hardware. In recent years, technical advances have made possible the rapid and efficient worldwide exchange of data. Today, rapid parallel data processing methods employing computer clusters and supercomputers enable large amounts of data to be analysed within an extremely short timeframe. In the commercial and scientific domains, special technological solutions, such as Apache Hadoop's MapReduce or Apache Spark (see below), make it possible to apply parallel algorithms to high-throughput computing. Parallel algorithms enable the simultaneous processing

11 Cf. Nakamoto 2009.
12 See https://deepmind.com/blog/trust-confidence-verifiable-data-audit [2017-10-17].

of numerous operations, significantly accelerating program execution.

Due to the major worldwide demand for the computing capacity required by big data and the resulting cost pressure, a widespread commercialisation of data infrastructure has taken place. The system architecture of a data centre today usually includes a large number of low-cost servers, which are networked both locally and with other server groups. Data is usually distributed over multiple servers and then processed so that transfer is minimised. This is a consequence of the technical characteristics of today's hardware, where computational power is often cheaper and thus more readily available than network capacity or the reading of massive datasets. Moving data is often more costly than processing it locally. Big-data applications are thus characterised by a distributed calculation extending across several networked computers; this calculation is highly complex, e.g. with regard to error tolerance and data security. It requires the development of an appropriately flexible software infrastructure that surpasses traditional database technologies. Many of the requisite techniques and components were originally developed by internet firms as internal tools.[13] Since about 2005, such software development has also increasingly taken place outside the major internet companies in the public or commercial domain. The Apache Foundation has exercised significant influence in this context with its numerous open-source projects (i.e. Hadoop, Cassandra, Spark or Kafka – programs totalling more than 120 million lines of code). In addition, there exist a large number of specialised start-ups offering their own software. The availability of such new big-data technologies has significantly reduced the technical hurdles to the development of corresponding services,

13 For example, the Apache Hadoop framework was developed as an open source implementation of Google's MapReduce programming model, and the Apache Cassandra database management system was originally created on Facebook.

which in turn acts as an economic stimulus.[14] The ways in which these techniques can be deployed not only within the technology sector, but also in the scientific domain and the public sector, including the healthcare sector, have since become apparent.[15]

Because of the increasing standardisation of hardware, software and protocols, the required computational power can be offered as a largely abstract service. The shift from locally administered computers to the less tangible, global virtual space of data centres is referred to as cloud computing. Cloud computing has become a highly competitive and very dynamically developing market, dominated in 2017 especially by Amazon Web Services, but in which other major IT and internet companies, such as Microsoft, Google or IBM, have expanded their presence.[16] The storage and management of data acts as an increasingly important cost factor in this market.

On the user side, it is above all start-ups and smaller companies who have benefited from these services, in that they can purchase computing capacity as needed. However, public institutions such as universities also often rely on such services for scientific applications. For their part, cloud providers are seeking to increase their share of the added value by means of a growing number of increasingly sophisticated services. Rather than manage a database on a computer they operate themselves, users often shift not only the computer administration to the cloud, but also the administration of the database itself. Should they require additional storage capacity or data throughput, either are available from the provider with the ease of a mouse click. Furthermore, the development of

14 According to estimates, the added value created by the Apache Foundation alone had reached approximately $10 billion by 2012 (cf. Greenstein/Nagle 2014, 624). This number is likely to be many times greater today.

15 For example, the Surveillance and Outbreak Response Management System (SORMAS), supported by public, especially German, research institutions, utilises a combination of an in-memory database and SAP cloud tools to provide a kind of early warning system for the risk of Ebola epidemics in West Africa (see Ensign et al., 2015).

16 Cf. Leong et al. 2017.

cloud solutions in the public domain is progressing at both the national and the European level, in the latter context as part of the EU strategy for high-performance computing.[17]

2.3 Data analysis and data science

Which ethical challenges arise from the use of big data in the healthcare sector depends primarily on the analytical methods that are employed to evaluate data, and the statements and decisions that are made on the basis of such evaluations. Relevant aspects include both classical methods of data analysis and their transformation in the age of big data, and new techniques in the field of machine learning and so-called artificial intelligence. Data science refers here not only to a new scientific paradigm of extracting knowledge from data, which has already begun to profoundly change scientific practice, but also to a new science of data that combines the methodical procedures of statistics and computer science, amalgamates, organises and annotates data and makes it available to collaborative analyses.[18] In view of the considerable increase in data collection, it is now possible to speak of a new type of data science, a so-called e-science (electronic science or enhanced science). E-science denotes the use of modern computer technologies to assist in preparing and performing scientific experiments, as well as in generating, sharing, distributing and retaining data.[19] New possibilities in data processing are fuelling data collection and the hope of gaining new insights and knowledge from this data. Due to the large volume and variety of data, scientific theories are increasingly being supplemented by automated analyses based on classical statistical methods as well as machine learning (see section 2.3.2).

17 See European Commission 2012a.
18 Cf., for example, Kitchin 2014 or Dijck 2016.
19 Cf. also Hey/Tansley/Tolle 2009.

2.3.1 Statistical modelling and validating correlations and mechanisms of action

An essential aspect of evaluating statements, predictions or conclusions based on data is the objectivity, reliability and validity of the data used and the methods of analysis employed. On this depends the legitimacy of explicit or implicit claims to truth, as well as the appropriateness of any resulting recommendations for action. Test theory and statistics – the science of working with data – are important tools in many sciences including medical research. In the era of big data, the need for the appropriate methodological competence has become obviously greater than before, due to the heightened complexity and heterogeneity of data, as well as the increasing possibilities for its linkage.

Many of today's big-data applications aim to identify patterns of potential causal mechanisms that can be used to explain empirical associations. These sorts of associations, such as those pertaining between the presentation of certain symptoms and the existence of a disease, are commonly referred to as correlations. The discovery and quantification of correlations is only the first step, however, forming the basis upon which assumptions regarding causal relationships within the determined context are made possible. One particular strength of big-data applications is that these assumptions can be empirically verified through further analysis. If (presumably) causal variables can be influenced, deducing possible interventions and assessing their potential effects constitutes another important application of big data (see below). Often the interest in collecting data is also guided by the goal of making statements regarding quantities that, either in principle or on pragmatic grounds, are not directly observable as such. The former includes, for example, statements about the future; the latter includes situations in which measurements are associated with high costs, high risks or side effects, or are prohibited for ethical or legal reasons.

The use of statistics is not limited to identifying empirical relationships between individual characteristics, on the basis of which correlations in an associated population can be established. A key task also consists of describing correlations on the basis of more complex models and thus mapping potential mechanisms of action. Against the background of such models, it is now possible not only to quantify correlations between variables, but also to describe them more precisely in terms of their direction of action and possible causal influencing factors. The ability to take into account additional variables that may condition or influence empirical relationships increases with the size of the sample. Because the significance of statistical tests is directly tied to the size of the sample of data collected, however, it must not be confused with the practical significance of any statistical relationships that may be discovered. Unlike correlation, statistical significance is not a measure of the strength of a connection. The extent to which any resulting statistical modelling of interrelations proves useful for predicting or planning effective interventions also depends on the number of additional variables that can be incorporated into this modelling. It also depends, for example, on the representativeness of the sample of data collected, the population under examination and the quality of the data collected. Only under the following conditions can statistical models hold up in the context of practical application: first, if the data of interest can be measured objectively, i.e. different observers, on the basis of the same information, arrive at the same estimates; second, if this data is sufficiently reliable, i.e. if repeated measurements yield comparable values, or if alternative indicators result in comparable estimates of characteristic values; and third, if it can be assumed that the data is valid, that is, at least one essential aspect of the characteristic under examination is actually being measured.

There are two ways in which analyses based on big data present opportunities: first, large datasets encompassing many particular cases are generated, allowing more precise

statements to be made. Second, various data sources and the diverse array of associated variables are linked with each other in such a way that previously unknown correlations can be uncovered. Diet apps, for example, collect data on eating habits that can be linked with patient data (e.g. pulse, blood pressure, blood sugar level, weight) to form datasets that, in traditional medical research, would typically only be attainable with great effort and a small number of cases. The relationships newly uncovered by big-data analysis can be used both in the context of predictions and in the context of interventions, and present the opportunity to empirically substantiate diagnostic, therapeutic or preventive measures. Evidence that data that has been collected at an earlier point in time (such as the daily number of steps) can significantly assist in the prognosis of later health (such as cardiac health) may suggest interventions that attempt to target and alter previously measured variables, on the presumption that they will influence subsequent developments. The availability of detailed information on the relationships between different variables, for which measurements taken at different times exist, thus contributes to the development and substantiation of preventive measures and efforts at therapy.

However, whether big data's strength and potential can be realised is not simply a result of the multitude and variety of available information. Rather, big data makes high demands on the integration of data in the context of statistical modelling. In the first instance, this has to do with the fact that as the number of variables and measurements increases, so too does the probability of at least partially contradictory information. Thus, the quality and significance of data (objectivity, reliability, validity) must be reflected upon, weighting must be applied as appropriate, and the opportunities and risks that may arise from the consideration of alternative variables in statistical models must be taken into account. Because big-data applications often rely on very different collection methods, the specific characteristics of these methods must be accounted for. It

should be asked to what extent different measurements can be regarded as equivalent or to what extent the method employed can have characteristic effects on the object of interest, which – as indicated – can in many cases not be directly observed. This raises the question as to which variables should be linked in the context of statistical modelling. Another problem that arises from the large variety of available information is that the use of alternative models can in principle allow for very different assumptions to be made. Alternative models arise, for example, from the fact that certain boundary conditions and parameters of a model can be defined differently. For this reason, the task of searching for possible falsifications, rather than simply confirmatory information, is important when developing models.

The independent review and verification of the results of data analysis is a central element not simply in the context of big data, but in scientific research and progress more generally. Such reviews can proceed by means of independent data sets (e.g. genetic studies in two independent samples, to achieve replication), empirical approaches that experimentally support or disprove a prediction, or finally, the use of alternative models (a scientifically provable statement should be relatively robust in the face of various models). In cross-validation, for example, a section of the data is randomly selected to be excluded from use in the model estimation, so that it may be used later to check the quality of the model. Another strategy for review is to organisationally or institutionally separate the processes of model generation and model validation. This is an important aspect of external quality assurance and verification; for scientific studies in the field of genetics dealing with large cohorts, it is often treated by leading journals as a requirement for publication. Methodological competence is thus a necessary but not sufficient condition for high-quality big-data analysis; the latter also requires the willingness to design algorithms without a preconceived result. Especially in the context of big data, the selective choosing and linking of data can have a manipulative effect, something which is difficult to

verify externally. The potential benefits of big data depend on the expertise and integrity of the individuals or institutions engaged in generating, selecting, linking and interpreting data.

Even if a model's predictions are proven correct, this alone does not allow for any statements regarding the correctness of the model used, or concerning a relationship of cause and effect. At stake is usually not so much the verification or falsification of causal statements, but the reliability of predictions. Furthermore, statistics by its very nature does not offer certainty for individual cases; it can only have a predictive value, which is accompanied by a certain probability of error. Probability is to be understood here in the sense of relative frequency and refers to repeated observation (of the same person or different people). A prediction regarding the likelihood of a particular outcome, such as a specific course of a disease, is thus a statement about what is to be expected according to empirical values averaged over a population of similar units, such as patients. However, due to the use of new machine-learning methods, such as deep learning (see section 2.3.2), it is increasingly possible to arrive at statements with a very high probability, which may also be of great relevance to the individual patient.

One cannot determine causal effects or mechanisms simply from correlations between variables. The relationship between two variables A and B, observed at a specific point in time, can be the result of the fact (a) that A causally influences B, (b) that B causally influences A, (c) that A and B affect each other, or (d) that A and B have a common cause. Furthermore, the observed relationship between A and B can ultimately result from specific features of the sample under consideration. It is sometimes difficult to distinguish between these different possible interpretations and further analytical steps are needed.

The value of data is often measured by the value of the statistical modelling it enables. Statistical models are, however, limited in one key respect: they do not allow conclusions to be drawn regarding the expression of relationships under altered conditions. Such changes may arise, for example, if the

underlying population changes. If one takes a model based on data collected in a specific environment – e.g. patients in a given geographical area or users of a particular app – the question arises as to how the correlations, and thus the model's predictions, change when looking at a different population – e.g. patients from another geographical area or non-users of the app in question.

Traditional statistical analysis places great emphasis on rigorously controlling the conditions of data collection, for example by avoiding unwanted distortions in sampling or by excluding anticipated disturbance factors. In the world of big data, however, this level of control, or even knowledge regarding the mechanisms of data collection, is often not possible or is only insufficiently available. This presents fundamental methodological challenges to the modelling of data and signals a need for caution in interpreting results and arriving at any (allocation) decisions with respect to individuals on the basis of such results.

Important and sometimes major changes can result from unknown or undetected factors including, for example, targeted interventions such as treatment by a doctor or lifestyle changes based on the tips provided by a health app. At issue here are no longer simply observed data and correlations, but questions of cause and effect – in other words, how such interventions may alter correlations and predictions. For example, the correlation between body temperature and symptoms may remain constant following the administration of an antipyretic medication, if the fever is the cause of febrile seizures and the drug thus reduces not only the fever but also the risk of convulsions. If, on the other hand, the fever is only symptomatic of another underlying disease, whose symptoms are not affected by the decrease in temperature, the correlation between body temperature and symptoms changes.

Answering questions about the effectiveness of interventions thus requires causal models[20] and causal arguments,

20 See Pearl 2010.

which are of an essentially different nature from findings derived from mere correlations. Neither the strength of a correlation nor the size of the dataset allows one to avoid this discussion. It would be a mistake to believe that more data automatically leads to more knowledge about causal effects. It would also be a categorical error to confuse correlational statements with causal statements.[21] The statistical proof of an empirical relationship and the determination of the effect size are independent from a deeper understanding of the mechanisms of action that bring about the effect.[22] Reasonable statements regarding cause-and-effect relationships ("causal effects") can be made even if the mechanisms of action are not or are only incompletely understood. Because of the complexity of biological systems, this is often the case in the life sciences. Medical progress benefits in many ways from an improved understanding of causal relationships, such as metabolic pathways in cells or regulatory circuits in organisms, but is not necessarily bound to such knowledge. Thus, we have known about the pharmacological efficacy of acetylsalicylic acid for more than 2000 years, while its first molecular mechanism of action (of many!) was only described in the 1970s. Even if science erred in its explanations, this would not contradict the validity of a causal effect.[23]

This distinction between causal effects and causal mechanisms – between knowledge of effectiveness and the "how" of

21 It must nonetheless be taken into account that, as yet, there exists in the philosophy of science no canonical reconstruction of the concept of causality; presumably, it is impossible to avoid pragmatic interpretations when reconstructing causal statements (for example, by linking the reliability of causal statements to experimental acts of intervention, unreal conditionals, or frameworks of rational betting behaviour). See, inter alia, Stegmüller 1983, 501 ff.

22 Nonetheless, a (pre-)understanding of mechanisms often leads to implicit or explicit assumptions that constitute a causal model.

23 Undoubtedly, however, the value of pre-scientific, operatively based insights can consist in their high degree of general applicability. Already in antiquity, this realisation was contained in the definition of *techne* or *ars* (art) as distinct from *episteme* or *scientia* (science). In this sense, medical disciplines are first and foremost "arts" of healing. See also Gethmann 1996.

the effect – remains pertinent in the face of big data. Big-data analyses can determine correlations that have a certain predictive power, independent of any interventions. If predictions regarding interventions are required, however, it is necessary to identify relevant causal effects and assessing their strength by means of additional arguments and assumptions or by obtaining additional data, for example from long-term or experimental studies. Correlative statements can also help to exclude certain relationships. Ultimately, one will attempt in the long term to develop a fuller understanding of the underlying mechanisms in an independent manner.

Above all, through new ways of accessing a much greater volume and variety of data, big data opens up the potential in medical research for the accelerated and improved discovery of new correlations between significantly more factors, as well as for the development of new hypotheses about their effects. Since the significance of statistical tests, as already mentioned, is directly dependent upon sample size, the question of the practical relevance of correlations must always be raised. This applies in particular to the interplay of the modes of action of weak and/or interacting individual factors which may have an effect in certain combinations. For example, it may be of interest to investigate which genetic factors are associated with an increased risk of cardiovascular or neurodegenerative diseases under which environmental and living conditions. These types of questions are explored in large-cohort studies involving thousands and sometimes tens of thousands of volunteers in order to identify indicators that can be used in early disease detection, prophylaxis and/or potential therapies.

2.3.2 Machine learning and machine perception

Machine learning is a rapidly growing field of computer science dealing with models and methods of data analysis that ties into numerous areas of application. Its historical roots lie

in disciplines such as statistics, physics, mathematics, pattern recognition and neuroscience. Usually, a given statistical model is "trained" with data. In the typical case of so-called automatic classification, systems use specific training datasets to "learn" rules of calculation that classify or categorise data in a certain way. The goal is always first to extract relevant features from raw data and then to contrast these features with regard to a particular problem or question. Apart from pure classification, there also exist techniques for calculating the probability of every alternative outcome or explanation or for making quantitative predictions. This abstract process has a tremendous and constantly increasing scope of application, including in the medical field.[24] For example, machines can learn to classify MRI images with regard to the presence and type of tumours or to develop highly complex decision-making rules and strategies that enable automated image processing in radiology.[25] From unstructured data such as patient voice recordings, it is also possible to glean information concerning a variety of physical and psychological features and disorders, ranging from fatigue and drug use to depression, post-traumatic stress disorder, psychosis and neurodegenerative diseases.[26]

It also holds true for machine learning that the larger the number of examples contained in the training dataset, the higher the predictive power of the model. Collecting data and generating and allocating labels or tags that code for the desired classification are important aspects of this process. Through its ownership of Instagram, for example, Facebook is in possession of one of the world's largest collections of photos in

24 Cf. Litjens et al. 2017.
25 Cf. Müller/Hanbury 2016.
26 On fatigue, see e.g. Krajewski et al. 2014; on drug use see Bedi et al. 2014; on depression see Scherer et al. 2013; on post-traumatic stress disorder and psychosis see Bedi et al. 2015 and Mota et al. 2012; on Parkinson's disease see Tsanas et al. 2012.

which humans are depicted[27], which can largely be identified by means of annotations or tags. It is therefore no surprise that Facebook has been able to develop facial recognition software of unparalleled efficiency[28], software that has also been used by companies in their applicant selection process. The British retail chain Tesco also employs facial recognition software to scan the faces of customers as they wait at the checkout, in order to optimise its advertising.[29] Diagnostic algorithms, developed with the help of machine learning and massive image databanks, are also growing in importance in the healthcare sector, for example, using photos to diagnose skin cancer[30] or predicting depression on the basis of image features and metadata extracted from Instagram posts.[31]

Besides the number of cases or training data, the breadth and depth of observation is also of essential importance to the quality of predictions generated by machine-learning models. The more measurements, features, and attributes are incorporated into such predictions, and the stronger the statistical relationship with the dependent variable, the greater the chance that a model will avoid errors caused by oversimplification. These basic statistical truths explain the effort in practice to link as many data sources to one another as possible. The importance of such data-driven predictive techniques in generating value on the internet cannot be overstated. What is new here is less their use in principle than, above all, their use to optimise micro-decisions in real time, something only made possible by the development of powerful computer and

27 Already in 2014, Facebook had a total of 250 billion photos (see http://www.businessinsider.com/facebook-350-million-photos-each-day-2013-9 [2017-10-17]). In addition, almost 450 million new images are uploaded to Facebook and Instagram together every day (see https://www.omnicoreagency.com/facebook-statistics [2017-10-17] and https://www.omnicoreagency.com/instagram-statistics [2017-10-17]).

28 Cf. Taigman et al. 2014, Taigman 2014 and Bennett 2017.

29 See https://www.theguardian.com/business/2013/nov/03/privacy-tesco-scan-customers-faces [2017-10-17].

30 Cf. Esteva et al. 2017.

31 Cf. Teramoto et al. 2017.

data-transfer structures, as well as the sheer volume of processed data. While many studies are based on datasets of between under a hundred and a few thousand data points, at issue here are datasets of hundreds of millions of data points, which grow every second with the use of internet services.

Since about 2010, major advances have been made in one branch of machine learning, so-called deep learning. Here, processes are modelled along a chain of hierarchically organised layers, similar to the manner in which neural networks operate in the human brain. Every layer thus relies upon the result of the previous layer while processing it further to generate new results. Because of the large volumes of processed data and increased computing power, the networks used today contain more intermediate layers than ever before, each building upon the work of the previous layer. The key difference to traditional processes consists of the fact that in deep learning feature extraction has been largely automated and models can often train directly on raw data. This reduces dependency on complex pre-processing steps, which can also incorporate the prejudices and mere intuitions of their developers.[32] For example, image recognition can begin directly on the level of the individual pixel or speech recognition directly on the basis of sentences as sequences of words or even letters, allowing for the recognition of natural speech.[33]

These effects can be achieved by models that recursively combine simple computational elements. Successively improved data representations are learned, whereby each representation is based on the previous one. In computer vision, for example, such models first learn local contrast filters, then simple visual features such as contours or corners, finally visual components (the wheel of a car, a human eye, etc.), culminating in a compositional representation of the subject matter and

32 However, prejudices can also be reinforced by the use of false or undesirable features of learning algorithms. Cf. Spielkamp 2017.

33 Cf. LeCun/Bengio/Hinton 2015.

scenes at hand. Similarly, speech recognition, signal processing, robotics and many other areas of application are already being shaped by deep learning.

Deep learning is often associated with the idea of artificial intelligence. Indeed, breakthroughs in the processing of sensory (especially visual and auditory) data have led to machines being able, in their own way, to "see" and "hear". This is referred to as a revolution in machine perception, and it constitutes a major step towards the development of robots that would be able to navigate our world "autonomously" and interact with people in a manner perceived as natural, for example, in the domain of home care. Furthermore, such systems would be suited to a wide range of other applications, such as evaluating the results of medical imaging techniques. The visual skills that radiologists, surgeons, laboratory assistants, etc., have acquired through training and expertise could be complemented or even replaced by automatic procedures. Very encouraging results have already been achieved in differentiating various traumatic lesions of the brain in MRI scans[34] or in classifying forms of lung cancer.[35] The auditory capabilities underlying the best speech-based interfaces (such as Apple's Siri, Amazon's Alexa, or Google Assistant) or the diagnosis of health disorders through automated voice analysis techniques, as mentioned above, have also already achieved a similar level of sophistication.

A key question concerns the extent to which such technologies will lead to the development of machine agents that possess the ability and authorisation to make decisions, and which could be involved in designing therapies or shaping health policy, for example by determining whether a new treatment is to be covered by health insurance. A major part of the internet is already automated, entailing interaction with intelligent systems that often remain hidden. The revolution of machine

34 Cf. Kamnitsas et al. 2017.
35 Cf. Teramoto et al. 2017.

perception outlined above, however, will open the door to the increasing application of these technologies in the real world. Thus, the penetration of everyday life by machines (pervasive computing) takes on a wholly new dimension: at stake is no longer simply the passive collection of data by means of sensors, but also active intervention in the flow of events by means of perceptive systems.

Alongside those machine-learning processes that rely upon the recognition of certain features, there are also important variants that function without feature extraction. The prime example is so-called recommender systems. The key concept underlying such systems is relevance optimisation, i.e. selecting the most essential from a plethora of available content, information and products. This requires self-learning systems that can identify relevant factors at the population level, i.e. factors that can be generalised across a large group of people, and then locate individuals and pieces of content within this system of coordinates. Factors corresponding to a user's particular viewing and consumer preferences, for example, can be derived from Netflix usage data. As a result, any individual can be grasped as a specifically weighted combination of precisely these interests within a system of coordinates plotted using such factors. The individual is thus situated in a space derived of features extracted from bodies of data, which as a rule can only be interpreted in a limited way, for it has not been conceptualised by means of existing, given dimensions. All that is needed in order to categorise the individual is their data. In order to identify the above-mentioned factors statistically, however, a large, collective dataset containing as many people as possible is required. Such recommender systems are already being developed for the healthcare sector.[36]

Much of the data collected is fed directly back into a service's data cycle, in order to optimise, adapt, or personalise services, often to the benefit of the user. The quality of search

36 Cf., for example, Bocanegra et al. 2017 and Gräßer et al. 2017.

results depends largely on usage data: what queries are submitted and how are they modified? Which results are clicked on and which are ignored? What role does the user's context or their current location play? If the available breadth or extent of this data is not sufficient, the relevance of search results and individualised recommendations can noticeably deteriorate. However, the ease and convenience entailed by the search for individually relevant information and content also necessarily entails the disclosure of personal data.

As a result of the developments in technological capacities mentioned above, which increasingly enable data analysis without perceptible delay, computers can respond ever more convincingly to users in dialogue form. This can potentially lead to the development of a "counterpart" who interacts with the user in real time and with whom the user can communicate as if from person to person. The central component of such a system is speech recognition, which breaks through the barrier between human and machine. Software-supported systems have already been in widespread use for several years, for example in the form of chatbots accessed via the internet, such as those used by transport companies to disseminate timetable information, or as preinstalled systems on smartphones or PCs, such as Apple's Siri or Microsoft's Cortana.[37] Since 2016, major providers have also sold bot systems based on standalone hardware, such as Amazon's Echo (Alexa) or Google Home (Google Assistant). These serve as distribution systems, even in areas pertaining to healthcare.[38] Such a device can also function as an "intelligent personal assistant" (IPA). On the one hand, such applications hold great potential in facilitating barrier-free lifestyles, particularly for people with limited

37 Chatbots are automated computer programs that allow users to simulate human dialogues via speech and/or text input and output.

38 Alexa is already being used as a sales channel for nursing and supplementary health insurance via Amazon's Echo loudspeaker (see https://www.versicherungsbote.de/id/4854127/Amazon-Versicherung-DFV/ [2017-10-17]).

mobility or impaired cognition. On the other hand, there is the risk of abuse, especially since chatbots are already so technically advanced that it is hardly possible for people interacting with them to determine whether they are communicating with a human partner or with software. Thus, deception and even manipulation of personal decisions are made possible.

2.3.3 Stratification and individualisation

Machine-learning methods in particular have the potential to identify differences within groups of people that were previously regarded as homogeneous. If in the past patients had to be stratified into a few main groups because, on the basis of relatively little data, only broad effects could be traced, big data in the medical field promises optimised diagnostics, prognosis and therapy that take better account of the specific conditions of the individual (so-called precision or "personalised" medicine). This brings about a new perspective on fundamental concepts such as disease: in different patients, the "same" disease is in fact not always the same, and is to be treated not in the same way, but individually.

By their very nature, stratifications always entail reductions of complex sets of individual characteristics to certain, single aspects. The fewer the groups delineated by this procedure, the more drastic the reduction of individuality and the greater the risk of erroneous classification. Especially in medical and scientific contexts, however, often only a binary grouping is possible, as, for example, in the decision for or against admission to a study cohort for a new drug. In other contexts, such as insurance, gradual stratifications can also be made, as in assigning policyholders to premium groups based on their age or certain other health-related data.

The formation of such groups, and the prognoses generated by algorithms as a result, may prove helpful to both the user of such analyses and to the people they affect. Their potential

margin of error can be made appropriate to the purposes for which they are employed. However, algorithms can also lead to problematic instances of erroneous classification, and complex big-data algorithms can make it difficult or even impossible to identify and eliminate the sources of error, depending on how, for example, the algorithm takes into account the occurrence and handling of certain sources of error or how certain boundary conditions are defined for a model. Thus, certain decisions regarding therapy for sufferers of rare diseases, whose characteristics diverge from the standard cohorts or strata, could be problematic. For example, while therapy with iron supplements would be helpful, or at least not harmful, for the vast majority of people diagnosed as anaemic by means of automated haematological analysis, in the case of hereditary hemochromatosis, which affects about one in three hundred Europeans, such a treatment would massively aggravate the condition. Such information must be incorporated into the algorithm from the start or taken into account and corrected retroactively.

2.4 The personal and health-related nature of data

From the wealth of information that can be collected and processed using the big-data methods described above, data with the potential to generate new insights in biomedical research, diagnostics, prediction, and therapy will be of most interest to the healthcare sector. Such personal health data is considered particularly sensitive because it offers deep insight into a very intimate area (see chapters 3 and 4). A key feature of current technological developments is that personal data can be collected and linked from a growing number of sources, and that this data can acquire, in the course of analysis, a relevance to health that was initially not apparent or expected (see section 2.4.3).

2.4.1 Data of a personal nature

Nearly all the data collected through the internet or by means of personal devices has either an explicit or implicit personal connection. This is a result, on the one hand, of the way this data is collected; on the other hand, it is essential to the way this data creates value. Only with reference to a person and their data, as well as the associated potential for data linkage, can value be created that can be monetised or clinically operationalised. The strength of the personal connection determines both the extent of this value and the reliability of any anonymisation measures.

Full identification generally proceeds by means of a person's name, especially to the extent that its unambiguous connection to the individual can be established on the basis of a mailing address, telephone number, e-mail address, credit card number, etc. Conventionally, pseudonyms are used as an alternative to the actual name when a person's identity is to remain hidden. Online, pseudonyms are often coded as user names or user identifiers. Especially of note is the use of so-called single sign-on (SSO) services. From the user's perspective, this is often a matter of using a variety of services, apps and websites without having to create a separate login for each one. Instead, many services allow for the (re)use of login data from Facebook (Facebook Connect), Twitter, Google, and other platforms. This leads to a considerable diffusion of data, by which the website in question gains access to a user's profile data, while at the same time the SSO service provider collects data on the user's web activity. If social log-ins were originally intended for communication between acquaintances and friends, they become in this way an instrument of data-collection in the hands of corporations, one that reduces the fragmentation of personal data on the internet in order to grant wider access to user behaviour and preferences. The conditional protection of privacy offered to the individual by data fragmentation thus tends to be eroded.

The personal nature of much data is only indirect, insofar as the underlying identifier (ID) is linked with a device (device ID) or a program (such as an internet browser). Especially relevant here are the device identifiers of smartphones, because as a rule these are used by just one person. These identifiers are administered by the manufacturers or platform operators, whose policies also regulate the further use of this data. The use of HTTP cookies, which store data in an internet browser and are thus able to unambiguously mark this data (so-called tracking cookies), is widespread. Additionally, cookie-syncing technology allows various services' IDs to be automatically synchronised, essential for the creation of comprehensive and persistent user profiles. This technology is particularly widespread in the online advertising industry (see section 2.5.4). Other technologies, such as canvas fingerprinting, are able to identify users without the use of any cookies; here, browsers are identified on the basis of characteristics such as installed plug-ins, fonts, versions, etc.[39] Today, a complex and largely inscrutable situation has evolved in which new ways are constantly being found to mark devices or browsers with a digital fingerprint and thus render them reidentifiable.

These sorts of identifiers are becoming increasingly relevant in situations pertaining to health and healthcare, for example through the use of mobile devices and apps that record the activities and vital signs of individual users (see section 2.5.5). This occurs in the context of internet searches or online purchases related to health-specific topics, or even with the aim of accurately documenting the use and billing of healthcare services.[40] The question of identifiers is so important because it not only determines whether privacy is preserved or violated, but it also represents the key to linking and accumulating data. If identifiers can be matched with one another, and at least one identifier can be unambiguously linked to a

39 Cf. Acar et al. 2014.
40 Cf. https://www.google.com/patents/US9740823 [2017-10-17].

person, anonymous data can be deanonymised. Similarly, simply through the matching of identifiers, surveillance can reach a level of precision such that there exists only one person to whom all data corresponds. For example, only four low-resolution space-time coordinates suffice to identify someone with 95 percent certainty.[41] This also applies to queries entered into search engines and online map services. Google Maps, for instance, can determine a user's home address to a high degree of probability, as it is often the destination when planning a route. Combining this data with search queries makes it easy to link this street address with an individual.

Alongside these implicit identifiers, there are also kinds of data that act explicitly as unambiguous marks of identification: genetic data, diverse biometric data (such as fingerprints or iris scans), or photos or videos of faces, which, for example, can be aligned with publicly available sources of data such as Facebook profiles. The controversy unleashed by the automated facial recognition and personal identification systems offered by Facebook, which rely on photos uploaded to the platform, clearly shows the social relevance of such questions of reidentifiability.[42]

2.4.2 Data of a health-related nature

Health-related data accumulate in various, sometimes overlapping, contexts, from medical practice and health-related research to government agencies and insurers to active and unintentional data generation by citizens or healthcare providers.[43]

In research and medicine (see sections 2.5.1 and 2.5.2), the use of increasingly powerful imaging and molecular biological techniques generates especially large volumes of data. Thus,

41 Cf. Montjoye et al. 2013.
42 Cf., for example, http://www.zeit.de/digital/datenschutz/2012-09/facebook-gesichtserkennung-dpc [2017-10-17].
43 Cf. Langkafel 2014, 14.

the size of the images produced by each digital X-ray, ultrasound, CT or MRI scan is typically measured in the range of megapixels or megavoxels. High-throughput screening procedures, such as the so-called omic technologies, also contribute to a massive increase in data volumes. The omic technologies include, for example, genomics (the study of the structure of genomes and the interactions between genes), proteomics (the study of proteins), metabolomics (the study of metabolism), as well as nutriomics (the study of the interaction of nutrients with the organism) and nutrigenetics (the study of the interaction between diet and genetics).[44]

Additionally, in the clinical context, medical histories, laboratory and imaging results, and diagnostic codes for administrative and billing purposes have been systematically collected for decades on everyone treated within the healthcare system. This data is, as a matter of principle, strictly protected by professional medical confidentiality; nonetheless it represents a resource of very high monetary value for big-data applications. The E-Health-Gesetz (E-Health Act)[45] envisages a tight schedule for the further digitisation of health data, setting 2019 as the date for the, initially voluntary, introduction of electronic health records. Private companies are already offering the option of mobile health records to help patients communicate with physicians and other healthcare providers.[46] In addition, there exists a wide variety of other health-related data that citizens or patients themselves record, administer, and share online using mobile devices and apps (see section 2.5.5).

44 Cf. Wirth 2015.

45 Gesetz für sichere digitale Kommunikation und Anwendungen im Gesundheitswesen sowie zur Änderung weiterer Gesetze of 21 December 2015 (BGBl. I, 2408). The law calls for the acceleration of the introduction and expansion of digital applications in the healthcare sector, such as the introduction of the electronic health card or the gradual replacement of previously paper-based processes with IT-supported procedures.

46 For example, lifetime.eu or medis-app.de.

2.4.3 Decontextualisation and recontextualisation

The technical and societal developments outlined above have far-reaching implications for the health sector. Traditionally, individual health data has been characterised by a particular sensitivity, due to its intimate connection to the personal sphere and the potentially negative impact of its being made public in social contexts such as the workplace (see section 2.5.3). This sensitivity must also be considered in the context of the legal and ethical evaluation of big data (see chapters 3 and 4). However, big-data technologies also allow for a comprehensive decontextualisation and recontextualisation of data that is collected, analysed, and linked for various purposes. This blurs, or at least is likely to blur, the boundaries between information that is considered health-related and information that is not, a distinction that is becoming increasingly difficult to maintain in the face of the accelerating linkage between all the areas of life in which data is or could be collected.

Individual datasets from disparate sources are first processed in a depersonalised and anonymised form; they are fed, so to speak, into an algorithmic "black box,"[47] in which this data is processed according to procedures, connections and rules which can hardly be understood from the outside.[48] These processes can produce different results: the original data is either grouped according to specific criteria, resulting in new clusters and divisions between cohorts, or transposed into a different context. In the healthcare sector, for example, such processes can link test results collected in the clinic to total genome analyses conducted in research laboratories, allowing improved treatment options to be selected for individual patients. In this way, they offer new ways to render diagnostics more precise and to establish promising therapeutic strategies.

47 On the concept of blackbox medicine see also Nicholson Price II, 2015.
48 See Bleicher 2017.

However, decontextualisation and recontextualisation also raise issues with regard to data protection and informational self-determination. They make the deanonymisation of data or the reidentification of individual users easier. Furthermore, if the probabilistic nature of the results of such data processing is ignored and people are categorised based on generalisations, premature or erroneous conclusions can be drawn regarding their characteristics, behaviours or even lifestyles. For example, if someone were to opt out of the bonus programme offered by their health insurer and regularly use an electronic loyalty card when buying wine in the supermarket, they may undeservedly be assigned to a high-risk group, despite the fact that they maintain an otherwise healthy lifestyle that is nonetheless not digitally documented, i.e. they exercise regularly without a fitness tracker or purchase healthy food at a farmers' market where no customer card is used.

Because all data collected in any form (for example, work and rest periods, places of residence and whereabouts, recreational activities, consumption patterns, etc.) can be interpreted in relation to personal health, it is in principle possible for all of this data to be categorised as health-relevant. Whether certain data should be regarded as sensitive or relevant to health, in view of these developments, can often no longer be determined at the time of its collection, but increasingly depends on the context in which it is used. This raises the question as to how a dynamic and context-dependent assessment of both the opportunities and risks of data use can succeed from a legal and ethical perspective. It also raises the question as to how, in this context, solutions can be formulated that can be put into practice in confronting the challenges posed by these opportunities and risks.

Willingness to share data in health-related contexts

Various studies indicate that the willingness to pass on one's data depends heavily on the context in which it is used. Initially, a key criterion for this willingness is the purpose for which data is collected and analysed. According to a Europe-wide study conducted by the Vodafone Institute, the majority of Europeans – especially in Germany – are critical of the mass collection and analysis of data.[49] However, scenarios in which anonymised health data in particular are collected and analysed, or passed on for research purposes, in order to improve the diagnosis and treatment of diseases are much more likely to be approved than the use of data derived from health apps in order to develop illness prevention strategies or even allow insurance companies to adapt their premiums. Nonetheless, according to a study by the polling agency YouGov, a third of users are willing to generate their own health and fitness-related data and share it with health insurers, provided that this brings them financial gain.[50]

According to a non-representative online survey conducted on behalf of the professional services firm PricewaterhouseCoopers, younger respondents in particular are prepared to provide private and, where applicable, family data to improve medical treatments.[51] Furthermore, it is relevant which health information is shared and to whom. Most acceptable is sharing data with doctors and clinics, followed by academic research institutes, health insurers, pharmaceutical companies conducting their own research, pharmacies, and finally government officials and neutral agencies such as consumer protection bureaux. In a representative survey conducted by the German IT industry group Bitkom, 60 percent of respondents would like to use an electronic medical record, in which all the data generated in medical practices, clinics or other healthcare facilities is compiled and stored electronically.[52] Of these respondents, three quarters would want to decide themselves which doctors were able to access to the digital information stored in their electronic records. Finally, a representative study carried out by the health insurance provider pronova BKK indicates that two out of three German citizens would release their personal data to be used by health insurers for consultation purposes, provided that all data is anonymised and processed in compliance with all applicable data protection rules.[53] Across all these studies, respondents' greatest concern was the loss of control entailed by the disclosure of data to third parties, especially if done for commercial purposes and threatening to bring about, because of the potential further use of data, possible disadvantages for the data provider.

49 Cf. Vodafone Institute for Society and Communications 2016, 15.

50 See https://yougov.de/loesungen/ueber-yougov/presse/presse-2015/pressemitteilung-self-tracking-rund-jeder-dritte-wurde-gesundheitsbezogene-daten-an-krankenversicherer-weitergeben [2017-10-17].

51 See https://www.pwc.de/de/gesundheitswesen-und-pharma/assets/personalisierte-medizin-studie-2016.pdf [2017-10-17].

52 See https://www.bitkom.org/Presse/Anhaenge-an-PIs/2017/03-Maerz/Verbraucherstudie-Telemedizin-2017-170327.pdf [2017-10-17].

53 See https://www.pronovabkk.de/downloads/14e3337132d6b4b5/Studie_Gesundheitsversorgung_2017.pdf [2017-10-17].

2.5 Actors and contexts of action

The collection, processing and use of large volumes of data involves various actors in diverse and sometimes overlapping functions and roles. The structures and motivations associated with each of these actors not only entail different parameters for data acquisition and evaluation; they can also influence in different ways the security, quality, and verifiability of the data collected and the ways in which this data is handled. Ambiguities arise not only from the multitude of different actors and the contexts in which they act, but also from their respective goals, assessments of opportunities and risks, and available resources. This multidimensional complexity suggests that no simple and coherent, all-encompassing normative judgement can be passed on the use of big data in the healthcare sector.

Thus, in what follows, a selection of contexts in which big data is employed will be examined in terms of the risks and opportunities they pose: first, biomedical research; second, healthcare provision; third, data use in the hands of insurers and employers; fourth, the commercial exploitation of health-relevant data; and fifth, the collection of data by affected individuals. Although these contexts will be dealt with separately below, they are interconnected and at times interpenetrating. The blurring of boundaries that is driven by big data, as described above, is also evident in the fact that research and health services, or commercial and non-commercial data use, ever more closely resemble each other and are increasingly difficult to distinguish.

2.5.1 Big data in biomedical research

The evaluation of large volumes of health-relevant data has long played an important role in scientific research. Large data volumes have held the promise of a substantial expansion of knowledge accompanied by increasing efficiency; not only

should they facilitate an improved understanding of relationships and processes relevant to biomedicine, but they should also improve preventive, diagnostic and therapeutic measures in medical practice. In the meanwhile, biomedical research has entered a new dimension of data and data integration, particularly in the field of genomic research, but also in neurological research. In the context of big data, researchers use and collect ever-larger quantities of data for their scientific projects, analyse it with increasingly advanced techniques, and link a growing variety of sources and types of data. This leads to a growing complexity in data analysis.

The central actors in the scientific domain include research institutes such as universities, non-university public and private institutions and their researchers and support staff, but also the subjects and patients who provide their data for research. In the research domain, large volumes of data are as a rule handled in accordance with high and easily monitored standards of data collection, use and security, particularly within research networks at the national, European and international levels that are especially geared towards working with large amounts of data (e.g. cancer registries, the Deutsches Forschungsnetz [German National Research and Education Network], PRACE networks). Geographically distributed scientific organisations are taking advantage of the new technical and infrastructural possibilities of big data and are networking for the purpose of data exchange and joint analysis and evaluation.[54] On top of this, scientific organisations are creating collaborative data infrastructures for long-term data retention, identification and replication, as well as semantic annotation

54 See, for example, the collaboration of the Indiana Biosciences Research Institute, the pharmaceutical company Eli Lilly and Company, the Roche Diagnostics Corporation, the Regenstrief Institute and the Indiana University School of Medicine to research the prevention and treatment of type 2 diabetes (cf. http://www.indianabiosciences.org/news/?newsId=39 [2017-10-17]).

and metadata searches.[55] Processes, software and standards from the commercial sector are also increasingly being employed (see section 2.5.4).

The ability to take account of diverse datasets in an integrative manner is seen as a particularly promising advance in the field of biomedical research, precisely because diseases are shaped by the combination and interaction of a number of factors, such as genetic predispositions, environmental influences or personal lifestyle. Because of this complexity, the probability, severity, course, and subjective experience of disease varies considerably from person to person, rendering prediction, prevention and diagnosis more difficult. Often, the underlying mechanisms of a disease are not yet or only partly understood, such that a targeted and thus potentially more efficient treatment is not possible. The ability to analyse large quantities of data holds considerable promise, for example in identifying features that are common to patients with a specific clinical presentation.[56]

In the healthcare sector, large collections of data are currently being compiled primarily from genomic data. Genetic factors play an important role in many health-related processes, and the technical possibilities for sequence analysis are already well advanced in this area. For example, in genome-wide association studies, researchers are investigating how often individual variants in the genome appear in combination with certain target features, such as particular diseases. With the help of such correlations, researchers hope to better assess genetically determined disease risks. However, the reliable identification of genetic features relevant to a given disease requires

55 See, for example, the "NCT DataThereHouse" project of the Nationales Centrums für Tumorerkrankungen Heidelberg (National Centre for Tumor Diseases Heidelberg) (cf. https://www.nct-heidelberg.de/das-nct/vorstellung/innovationen/nct-datatherehouse.html [2017-10-17]).

56 Concrete examples with a direct influence on the treatment of patients include the reclassification of mutated BRCA1 or BRCA2 genes from "disease-causing" to "neutral" and the reverse reclassification from "probably neutral" to "probably disease-causing" for certain mutations in the CTNS gene. See Lek et al. 2016.

a sufficiently large sample size. In the field of Alzheimer research, for example, cohorts of many tens of thousands of patients are being studied.

Medical research by the genetic testing company 23andMe

The inclusion of resources outside of the work of classical science can create new ways to acquire sufficiently large amounts of data. For example, the US biotechnology firm 23andMe, which markets genetic analysis directly to private individuals for the purpose of researching ancestry and assessing health risks, offers its customers the option of releasing their data for use in research projects in collaboration with public or private research organisations. For research into Parkinson's disease, for example, the company compared the specific gene regions of 3,400 Parkinson's patients, provided by patient organisations and clinics, with those of nearly 30,000 corporate clients who were unaffected by the disease, thus identifying two new gene variants that contribute to a hereditary predisposition for this disease.[57] 23andMe advertises that, by releasing their data, its customers can contribute to hundreds of studies, and thus play a significant role in the advancement of scientific progress.[58] However, the company has also been criticised: by treating the genetic and supplementary health data collected from its customers as company property, it could also ultimately hinder research or the use of research results, namely, if decisions regarding the future use and release of this information are made for primary financial reasons.[59]

Thanks to the greatly reduced cost of gene sequencing, more extensive genetic analyses with correspondingly larger data volumes are playing a growing role. The technical costs of complete genome sequencing already lie well under 1,000 euros and will continue to fall.[60] It is estimated that by 2025, 100 million to two billion human genomes will be sequenced.[61] These

57 Cf. Do et al. 2011.
58 Cf. https://www.23andme.com/en-int/research [2017-10-17].
59 Cf. Seife 2013.
60 At Veritas, for example, genome sequencing currently costs 999 US dollars (cf. https://www.veritasgenetics.com/mygenome [2017-08-30]). Illumina states that its devices will enable genome sequencing for 100 US dollars in the future (cf. https://www.illumina.com/company/news-center/press-releases/press-release-details.html?newsid=2236383 [2017-08-30]).
61 Cf. Stephens et al. 2015. For comparison: According to Francis deSouza, President of Illumina, a manufacturer of machines for genome sequencing, a total of about 1.6 million genomes will be sequenced by the end of 2017 (cf. Regalado 2017).

will include so-called exome analyses, in which not merely select gene variants, but all genes of the examined person are analysed, or even whole genome analyses, which also include genetic material that does not code for proteins. Studies of this kind represent greater opportunities to understand biological regulation mechanisms, such as, for example, those that could contribute to tumour development. For not only are individual molecular markers recorded, but so too are copy number changes – structural variations that can lead to the formation of fusion proteins – and changes in non-coding regions found specifically in the tumour.

New treatment options for brain tumours using total genome analysis

A study of children with a certain type of brain tumour, in which the entire genome was sequenced from, to date, 400 tumour specimens as well as from blood samples, revealed that almost all patients exhibited a change in a molecular signalling mechanism that is central to the growth control of cells and already scientifically well studied.[62] Thus a whole range of potential new medications for this type of tumour have become available, which could be particularly effective in combination, and which in the future could enable a more gentle treatment for affected children. The raw data has been stored in the European Genome-phenome Archive (EGA) and is available on request to all consortium members and patients. This example also shows that big-data analysis in medical research can quickly have an impact on medical practice (see section 2.5.2).

However, in the case of many diseases, the combination and interaction of factors is extremely complex and changes in the genome play only a limited role. To truly improve diagnosis, therapy and prevention over the long term, more extensive databases must be created that integrate different kinds of information. Big data has great potential in this regard, as it makes possible for the first time such comprehensive and cross-source analyses on a truly large scale. Genetic data could, for example, be supplemented by other omic data (such as data

62 Cf. Jones et al. 2013.

pertaining to the transcriptome, the proteome, the metabolome, see section 2.4.2), image data, or clinical data. Critical to this kind of integrative analysis is not only the sheer quantity of data fed into it but also the quality of its preparation and interpretation. The integrative analysis of a diverse range of health-related data with the help of systems biology methods, which are becoming increasingly powerful thanks to machine learning (see section 2.3.2), is making it possible to better understand biological and medical interrelationships, on the basis of which much more accurate diagnoses and prognoses can be made.

Major worldwide efforts are currently being made to create patient databases on an as-yet-unknown scale, by means of both public funding and massive private investment. The British UK Biobank, for example, has registered since 2007 approximately 500,000 participants and offers researchers the opportunity to use its data sets in pursuit of their own scientific enquiries.[63] The German "NAKO Gesundheitsstudie" ("NAKO Health Study", formerly "German National Cohort") has followed approximately 200,000 subjects since 2014, in order to study the causes of widespread diseases such as cancer, dementia, diabetes, cardiovascular diseases, and infectious diseases.[64] The EGA stores and administers omic data, especially genome data; currently, it contains more than 3,500 datasets from over 1,500 studies.[65] The Alzheimer's Disease Neuroimaging Initiative (ADNI)[66] and the study "1000 Gehirne" ("1000 BRAINS")[67] are giving researchers access to extensive data pertaining to the brain and its ageing and neurodegenerative processes.

63 See http://www.ukbiobank.ac.uk/participants [2017-11-17].
64 See http://nako.de/studienteilnehmer/das-untersuchungsprogramm [2017-11-17].
65 See https://ega-archive.org [2017-11-17].
66 See http://adni.loni.usc.edu [2017-11-17].
67 See http://www.fz-juelich.de/inm/inm-1/DE/Forschung/1000_Gehirne_Studie/1000_Gehirne_Studie_node.html [2017-11-17].

The goal of integrative data collection is to identify disease-relevant changes through large-scale comparisons. Because of the high variability and complexity in the relationships that contribute to the development of diseases, especially large patient datasets and suitable control groups are indispensable to valid analysis. The resulting need to pool data collected by multiple institutions in often disparate contexts poses particular challenges for the use of big data in medical research, especially regarding data standardisation and quality, data protection and data access.

While uniform data formats have long been successfully used in other sectors, for example the tourism industry, in the healthcare sector this has only been implemented in certain areas. One example is the international data standard HL7 (Health Level Seven), which is designed to provide specifications for the presentation of medical data and information and thus enable communication between institutions and various parts of the healthcare sector. This data standard is however used only within hospitals, but not for data exchange between the clinical and the primary care sector; instead, other data exchange formats are used in this area. Non-profit organisations, such as the Clinical Data Interchange Standards Consortium (CDISC), advocate the introduction of national and international standards for data sharing in clinical trials, and have already achieved partial success. For example, data standards developed by CDISC are accepted in the USA by the Food and Drug Administration (FDA) for the submission of clinical data for drug approval.

Nevertheless, in the healthcare sector, scientific and clinical data is still mainly collected according to very different programs, some of which are specific to the institute in question. Unifying interfaces and transferring data and any associated metadata would facilitate comparison and novel combinations of data, thus opening up new treatment options. Ideally, data should be combinable across various institutions, clinics, and research centres. However, this would require a coordination

of data formats and protocols. Furthermore, the quality of the data must be constantly monitored and its evaluation continuously improved. For this purpose, an exact documentation of the origin of the data, its annotation as well as the steps taken in its processing (provenance tracking) is of essential importance. In this context, graph-based metadatabases, so-called knowledge spaces, have been developed, which allow comprehensive semantic searches (in contrast to simple searches by keywords). This requires the careful curation of data, a demanding, labour-intensive process in which the nature of any given data must be accurately described. Thoroughly describing data in this way also fulfils one of the chief prerequisites for the reproducibility of analyses.[68]

Furthermore, it is important to clarify who may access data and under which conditions, in order to facilitate a transparent exchange while at the same time ensuring high data protection standards, especially in the case of large-scale collaborative projects. Clear guidelines are also necessary regarding the circumstances under which patients and subjects may access their data, how they may consent to the use of their data in a dynamic and modular way (see section 4.1.2), and how data will be preserved in the long term. In practice, well-functioning rules for data exchange are often lacking. This is partly due to privacy concerns. Much health data, e.g. neurological imaging data, is unique to each person. The possibility thus arises that datasets can be attributed to individuals. Furthermore, this possibility cannot be eliminated by techniques such as skull stripping, whereby brain images are separated from images of the surrounding bone, skin, and connective tissue, and thus from the face of the scanned individual. Concerns about the possibility of identification are one reason that such data is handled in an especially restrictive manner. Data exchange is also often made more difficult in practice by the fact that suitable channels of contact and communication and models for

68 Cf. Livingston et al. 2013 and Davis-Turak et al. 2017.

consenting to the secondary use of their data are not available to patients and test subjects.

The very restrictive handling of data that can result from such concerns and practical difficulties represents an obstacle to scientifically legitimate efforts to, for example, research rare diseases, or to identify weak but interacting factors for certain other diseases, with the help of big-data-driven analysis carried out across institutional, regional, or national borders.[69] Especially in the latter case, there are major bureaucratic hurdles to overcome. The EGA, mentioned above, not only contains a publicly accessible area, but also a restricted access area currently regulated by over 400 committees.[70] This generates significant administrative costs and delays for researchers interested in conducting further analysis with this data.

Possible solutions are offered by new models of consent to data use (see section 4.1.2) as well as technical processes that could allow an uncomplicated exchange of data while minimising the risk of identification. There exists, for example, the ability to analyse imaging data locally, to extract certain features from this data, and only to export the latter (in situ querying), while nonetheless combining these features with those extracted by other institutions.[71] It is also possible to work with group data or data that has been rendered artificially "noisy", such that tracing this data back to any individual would be extremely difficult. Nonetheless, the question of a given instrument's suitability depends on the application at hand. Under certain circumstances, one may, for example, even retroactively identify individuals if and because data indicates an acute disease or a life-threatening condition. The techniques mentioned above also require extensive methodological developments in databank management, annotation and metadata extraction. This in turn requires institutional support, such as,

69 Cf. Nationales Aktionsbündnis für Menschen mit Seltenen Erkrankungen 2016, 14 f.
70 Cf. https://www.ebi.ac.uk/ega/dacs [2017-11-07].
71 For example, Frackowiak/Ailamaki/Kherif 2016.

for example, the newly founded Center for IT Security, Privacy and Accountability (CICSA) of Saarland University intends to provide.

Data exchange is not only rendered more difficult by privacy concerns, but also by uncertainty and different ideas about who has the right to avail themselves of the generated data and to what extent. From the perspective of many researchers in laboratories and clinics, data "belongs" initially to whomever conducted or commissioned the relevant measurements, although this is legally incorrect (see section 3.2.2). As a result, it is often this person who wishes to decide to whom to grant access to this data. Institutions also frequently regard data as their "property", because, among other reasons, the generation of this data represents a considerable investment in both financial and human resources. Furthermore, the exclusive publication of studies based on the data in question is often associated with recognition for researchers and institutions, as well as the allocation of research funds. For firms, commercial motives also stand in the way of generous data sharing practices.

However, especially in the case of fully or partially publicly funded projects and data applications, such efforts to monopolise the use of data can lead to conflict. A significant proportion of research and especially of the research infrastructure is indeed publicly financed, and many citizens contribute with their data to building up these "data troves" in the first place. Furthermore, some wearable medical devices collect health data that their manufacturers, despite being financed through the social insurance system, refuse to make available for use in research.[72] Many researchers, referring to the clear public interest in sharing in the insights and profits derived from these data flows, thus demand a say in the use of data collected in such contexts.

72 In this context, the medical device manufacturer Medtronic, for example, has been criticised for not making the data obtained from its widespread blood glucose sensors and other devices available to external researchers and in some cases even to the users themselves (cf. Wilbanks/Topol 2016, 347).

These challenges can be met by various solutions. While some states already centralise much health data (e.g. Greece and several Scandinavian countries) and some institutions allow research data to be handled in a truly open way (e.g. the UK Biobank), Germany has thus far remained comparatively restrictive in terms of data exchange. The German E-Health Act, however, seeks to open up the telematics infrastructure for health research; this infrastructure is currently being developed to facilitate an improved and more secure networking of the actors in the healthcare sector. Likewise, a study commissioned by the Federal Ministry of Health on the further development of the German e-health strategy highlights the need for a closer integration of data collected in the healthcare system and biomedical research.[73] The Federal Ministry of Education and Research, furthermore, launched a medical informatics initiative in 2017, which aims to combine diverse datasets as well as different methods of storage and analysis in a single, national infrastructure.[74] This initiative initially focusses on university clinics and their partners, which already maintain close ties between healthcare provision and clinical research. Other research institutes, universities, private clinics and firms in relevant sectors such as IT, pharmaceuticals, biotechnology and medical technology are to be included as the initiative moves forward.

Non-profit initiatives also exist, in which citizens can autonomously feed their health data into databases for dissemination to researchers in compliance with security standards (and in combination with reliable opt-out mechanisms), with the goal of efficiently advancing the state of both global and individual health and healthcare.[75] Examples include the US open science organisation Sage Bionetworks[76], which has already developed a number of models for collaborative

73 Cf. Strategy&/PricewaterhouseCoopers 2016.
74 See http://www.medizininformatik-initiative.de [2017-10-17].
75 Cf. Wilbanks/Topol 2016.
76 See http://sagebase.org [2017-10-17].

research, and the Swiss organisation MIDATA[77], which offers citizens the opportunity to avail themselves, via non-profit co-operatives, of repositories for the secure storage, management and dissemination of various kinds of health-related and other personal data. Sometimes completely open access to data is possible. Open Targets[78], a joint initiative of publicly funded institutions and firms, aims to establish a platform to validate therapeutic targets through genome-wide studies and make the data freely available to the public.

2.5.2 Big data in healthcare

The healthcare sector is characterised by a variety of actors, some with divergent interests.[79] These include, for example, the providers, insurers and, of course, recipients of health services, but also authorities, interest groups and researchers with direct links to medical practice. Among healthcare providers, public and private clinics and their specialist and IT departments are particularly relevant in terms of big data. They carry out various functions related to the generation, use and administration of large volumes of data. The handling of large data volumes from a diverse range of sources (for example, gaining access to specialised databanks) is also becoming an increasingly relevant activity for practicing physicians and their staff, as well as pharmacists and other health service providers.[80] As a technical basis for facilitating data exchange, the E-Health Act

77 See https://www.midata.coop [2017-10-17].
78 See https://www.opentargets.org [2017-10-17].
79 Cf. Deutscher Ethikrat 2016, 8 f.
80 Via the West German Teleradiology Association under the umbrella of MedEconTelemedizin GmbH, for example, clinics and medical practices already exchange about 35,000 examination results per month on patients (see https://www.medecon-telemedizin.de/news/westdeutscher-teleradiologieverbund-nimmt-auf-conhit-den-300-teilnehmer-ins-visier [2017-10-17]). In the FALKO.NRW project, which was launched in 2016, data exchange is to be further improved and the networking of the approximately 300 participating facilities is to be intensified (see https://falko.nrw [2017-10-17]).

calls for the establishment of system-neutral, interoperable interfaces, which should also be available to industrial suppliers and scientific institutions. Providers of telemedicine services should also expressly be included (see Section 291d-e SGB V).

Other actors in the healthcare sector include the public law bodies by which the medical community governs itself, as well as associations formed under private law such as professional societies and interest groups, for example the Bundesverband Internetmedizin (Federal Association of Internet Medicine), founded in 2012. Within the scope of their legal responsibilities and/or voluntary remits, these actors are also entrusted with collating, administering, preparing and ensuring the accessibility of data. In addition, they increasingly assume responsibility for the further development of internet-based health services and the monitoring of these services in line with data protection standards. Patient organisations, on the other hand, are primarily concerned with protecting the interests of patients and ensuring their reliable care. They therefore have an interest in the best possible care, but also in data security, in anonymisation procedures that are as reliable as possible, and in the use of personal health data in such a way that allows for informed consent. To this is added the role of advising patients on medical services. Finally, there exist commercial services, such as online doctor rating sites, that occupy a grey area between independent patient advice and advertising by healthcare providers.

Finally, diverse actors from the areas of health policy, health administration, and other institutions concerned with public health collect and use data in massive quantities, for example, in order to generate prognoses for the long-term development of age-associated diseases or to anticipate the timing and intensity of a flu outbreak.[81] The organisations involved include ministries, public health departments and the Robert Koch Institute, whose public mandate is to monitor and prevent

81 Cf. http://www.flu-prediction.com [2017-10-17].

disease, but also other institutions such as the Bundeszentrale für gesundheitliche Aufklärung (Federal Centre for Health Education), the Gemeinsame Bundesausschuss (Federal Joint Committee), or the Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Institute for Quality and Efficiency in Health Care). The statistical offices of the German federal states, the Statistisches Bundesamt (Federal Statistical Office), and Eurostat (the statistical office of the European Union) collect and analyse data on a wide range of health topics and make the information available to other offices, policymakers, economic actors and the public.

Alongside the improvements in disease research that many hope big data will bring – more precise understanding of the origins of diseases and their appropriate treatments – the employment of big data also holds great promise for medical practice, particularly in the form of more personalised treatment concepts as well as improvements in efficacy and efficiency. Experts predict that algorithms will dramatically improve diagnostics in numerous domains and will become more accurate than doctors themselves.[82] Advances in artificial intelligence systems are opening up new opportunities, for example, in the (partly) automatic analysis of medical images. Medical imaging has become an important discipline for engineers, and advances in this field, including in real-time capabilities, promise to make innovative diagnostic and treatment procedures a reality – for example in computer- assisted surgery, by which an exact model of the affected region is generated on the basis of comprehensive image data and can be used when planning and navigating the operation. Likewise, new ways are being created to better monitor disease through, for example, telemedicine, to minimise the risks of harmful interactions between medications, and to simplify the transfer of patient information when changing doctors or receiving treatments

82 Cf. Obermeyer/Emanuel 2016.

from multiple healthcare providers. This is especially important in cases of complex or rare diseases.[83]

Being able to rely upon large volumes of relevant data also improves the stratification of patients, thus reducing the chances of side effects and unnecessary therapeutic interventions.[84] Connected to this is the hope of reducing costs for the healthcare system and accelerating the development of new therapies, e.g. in the pharmaceutical industry. Furthermore, patients should more quickly be able to benefit from therapeutic concepts that have already been successfully employed with other patients.[85] Optimised early recognition can improve the prognosis for a diverse range of diseases. Healthy lifestyles can be encouraged by the use of health apps and personal devices that collect data, even though there exists a potential for negative consequences, such as a fixation on the self-tracking of bodily states and unnecessary medical treatment (see section 2.5.5). The collection and evaluation of health-relevant data has an enormous preventative potential, to the extent that certain genetic carriers could be able to avoid group-specific risks. Furthermore, the potential held by big data, as outlined above, promises to enable the state to better plan for the demands that will be placed on its healthcare system, and thus to reduce its expenditures.

The opportunities presented by data-intensive approaches, however, are paralleled by risks. These lie, for example, in the actual or perceived loss of control over one's personal data and the continuous erosion of limits on the ability of service

83 More than 550 patients with tumour diseases have already been recruited at the National Centre for Tumour Diseases in Heidelberg. Within the framework of a joint evaluation of, for example, genome examinations and transcriptome analyses, relevant correlations between disease progression and genetic profiles will be uncovered and new and personalised therapy options developed (cf. Horak et al. 2017).

84 Cf. expert forum "Digitisation and Health" in the Hightech Forum 2017, 14.

85 Cf. the interview with Hasso Plattner in the special supplement to the Frankfurter Allgemeine Zeitung, 17 April 2015: http://www.anna-seidinger.com/pdf/FAZ_MedizinZwischenMoeglichkeitenUndErfolg_2015.pdf [17.10.2017].

providers to access one's intimate information (the "transparent patient"), as well as the concomitant increase in the potential for unauthorised third parties to misuse this data (insurers, employers, media, the latter particularly in the case of persons of public interest). To this is added the concern that an increased use of big-data-supported approaches could further reduce the personal attention paid towards patients, as doctors and patients will to a greater and greater degree communicate via machines rather than directly. Likewise, there arises the danger of misdiagnosis resulting from an uncritical or improper use of big-data-based recommendations.

Wherever the state has access to the relevant data, data exchange across the borders that separate authorities in the domains of, for example, health, social services, and labour could subject citizens to an unjustifiable regime of surveillance. Furthermore, a system of comprehensive, aggregated data pertaining to each citizen and their behaviour could even be used for "education" according to the ideas of a totalitarian state. A "social credit" system would then, for example, decide an individual's degree of access to public and private services – from transportation to education to credit and insurance. The first examples of such developments already exist.[86] It is obvious that such practices would be unconstitutional in this country, but even the mere existence of such possibilities is sufficient cause for concern.

2.5.3 The use of health-related data by insurers and employers

When assessing the risks associated with the future course of someone's health is key to a private, economic decision it is standard practice to determine that person's current state of health and medical history. This applies, for example, to

86 See, for example, Condliffe 2016 and Dorloff 2017.

certain contracts (life insurance, occupational disability insurance, private health insurance, credit agreements). Similarly, public authorities also assess the health of those applying for the status of civil servants. Statutory health insurance funds, on the other hand, are obliged to accept applicants without prior medical examination (obligation to contract); however, like private health insurers, they too have a certain interest in monitoring their policyholders, in order to maintain policyholders' health and reduce their own costs as much as possible by offering timely advice and incentives. For the same reason, employers, including the federal government, the federal states and municipalities may also be interested in obtaining health information about their employees or civil servants on their payrolls.

Here, big data opens up a wealth of new possibilities for accessing and evaluating relevant information that are not fully covered by applicable law (see chapter 3). Growing data volumes and proliferating data interconnections allow increasingly individualised profiles of individuals or groups of people to be generated. Such profiles orient themselves towards those properties that are considered meaningful for a particular purpose. Insurance contracts and private loans are among the cornerstones of individual livelihood. Outside of the solidarity-based system of public, statutory pension and health insurance, decisions regarding such contracts are overwhelmingly in the hands of private companies. These firms thus have a high degree of responsibility in dealing with the data required for evaluating and reaching such decisions.

Genetic diagnostics and risk prediction

In some cases, genetic information allows one to predict the course of a disease with a certain probability. This is a key reason why the use of such information in preparing employment or insurance contracts is subject to especially strict regulation in the form of the Gendiagnostikgesetz (Genetic Diagnosis Act, GenDG). The GenDG expressly prohibits insurers

from receiving or making use of data from genetic tests or analyses. A key purpose of this legislation is to provide protection against discrimination. Of course, data derived from genetic tests and behavioural data cannot be simply equated. However, the predictive and prognostic use of behavioural data can also lead to discrimination. In its 2007 Opinion "Predictive Health Information in the Conclusion of Insurance Contracts," the National Ethics Council (translator's note: precursor of the German Ethics Council) already pointed out that only regulating the use of predictive and prognostic health information of a genetic nature does not adequately address the ethical core of the problem.[87]

The GenDG has been in force since 2010, but in recent years an increasing number of clinical situations have arisen in which the disclosure of genetic characteristics to payers in the healthcare system may be in the immediate self-interest of the patient. In the case of hereditary breast and ovarian cancer, for example, this could facilitate access to specific preventive measures (magnetic resonance mammography, preventive mastectomy) or therapies (chemotherapy with PARP inhibitors).

Among the basic principles of insurance against the risks that arise over the course of one's life, and against which one often cannot alone prevail, is the distribution of risk across a community of insured. An insured event brings about a redistribution of resources from the collective of insured to the individual beneficiary. In this sense, the quantification of individual risk profiles prior to the insured event, and thus a fair structuring of premiums, ultimately serves not only the self-interest of the insurance company, but also social justice (see chapter 4). Refining the basis on which premiums are calculated can increase the profitability of the insurance industry and may ideally lead to greater cost-efficiency in the interest of those paying into the system. A similar principle applies to credit agreements: granting a loan only under the condition of sufficient creditworthiness protects the collective of borrowers from excessive default risks and, often, the individual applicant from being financially overstretched.

By its nature, big data is ideally suited to the task of generating risk scores from a complex network of individual factors;

87 See Nationaler Ethikrat 2007. For predictive health information on recruitment screening, see Nationaler Ethikrat 2005.

in the insurance industry, intensive efforts are already underway to implement big-data-driven risk assessment and assignment.[88] It is generally accepted that insurance companies are justified in structuring premiums, for example for life insurance policies, on the basis of individual data relevant to risk, such as previous illnesses. However, apart from such objectifiable "hard" criteria, which are usually made transparent to applicants who have been rejected or made to pay additional risk surcharges, contracts are also structured with reference to a number of "soft" assessment parameters. These include, for example, the stability of one's social environment, one's consumer habits, or the degree of hazard inherent to one's recreational activities. Of particular importance are individual health risks that may be of very real importance, but may also be difficult to quantify and verify, such as early tobacco use or compliance with certain preventive health recommendations. Not only private insurers, but also statutory health insurance providers are increasingly beginning to take such factors into account, for example, by offering bonuses to non-smokers or participants in prevention programmes, thus developing a practice of individualising risks.[89]

Insurance, by its nature, involves largely unforeseeable events. Thus, if big data were to predict, to an excess of detail, the life course and expectancy of the insured, this would call into question the business model of the private insurance industry. On the other hand, potential information asymmetries (in both directions) could give rise to opportunities for abuse in the form of discrimination and/or adverse selection. For example, the analysis of commercially available, big-data-generated personal behavioural profiles, perhaps even access to genomic data from scientific joint projects – which are not subject to the provisions of the Genetic Diagnosis Act – could

88 See, for example, Hauner 2016.

89 Cf. https://www.krankenkassen.de/gesetzliche-krankenkassen/leistungen-gesetzliche-krankenkassen/praevention-vorsorge-krankenkassen [2017-10-17].

enable insurers to specifically select low-risk applicants. Conversely, a potential policyholder could first secretly undergo genome analysis abroad, in order to assess their future risk of disease, and then tailor their own insurance cover on the basis of this information.[90]

Fundamentally, if big data brings qualitative improvements to risk stratification mechanisms, this can also increase the quality of overall risk protection for the collective of policyholders or borrowers. However, scenarios become problematic in which, given a constant budget, the bonus granted one insured person equals a penalty for another who does not receive such a bonus. Accordingly, insurers, including statutory insurers, have a legal as well as ethical obligation to justify such bonus programmes (see chapter 4).

Furthermore, the lack of transparency often associated with big data can cause problems by obscuring the factors that underlie qualitative or quantitative allocation decisions that immediately impact the individual, for example, the imposition of a risk surcharge and/or the rejection of an insurance or loan application. This is likely to make it difficult for those affected to dispute or take legal action against such disadvantageous decisions. Sometimes the erroneous interpretation of a single parameter can have serious consequences for the applicant or policyholder. While the insurance industry has already set up low-threshold complaint procedures[91] and is subject to oversight by the Bundesanstalt für Finanzdienstleistungsaufsicht (Federal Financial Supervisory Authority, BaFin), it must nonetheless be questioned whether the competence and capacity currently available is equal to the specific challenges presented by big data.

Decisions made by employers in processing applications are also of considerable importance to those affected, but the underlying parameters on which these decisions are based are

90 Cf. Henn 2016.
91 For example, https://www.versicherungsombudsmann.de [2017-10-17].

largely non-transparent. Even though including health examinations in the application process is usually prohibited, information available online concerning a given applicant is already often used in the official selection process. About half of all major companies already check any available social media profiles in order to gain a better impression of an applicant's character.[92] The ability to extract, using big-data-based analysis, genuinely health-related information from such data is constantly increasing, as is the number of commercial services offering such analytic tools.[93] At present, legal assessments of such approaches are divergent.[94]

Even within existing contracts, employers and health insurers have an interest in the health of their employees and policyholders, since cases of illness can generate high costs. The monitoring of employee or patient behaviour opens the door to incentives for healthy, and sanctions for unhealthy lifestyles. These need not take the form of financial bonuses or penalties. They may also consist in the collection of information or in the promotion of "healthy" behaviours, when, for example, devices or apps are provided to measure health-related factors, or when people are motivated and facilitated in pursuing a relatively active lifestyle. Manufacturers of wearable devices are already cooperating with employers and insurers.[95]

Such programmes open up opportunities for everyone concerned. Lower rates of illness benefit not only employers and insurers, for whom costs are reduced, but also workers and policyholders, whose health is improved through such preventive

92 Cf. https://www.bitkom.org/Presse/Presseinformation/Jedes-zweite-Unternehmen-ueberprueft-Bewerber-in-Sozialen-Netzwerken.html [2017-10-17]; Weitzel et al. 2016.

93 Cf. https://business.linkedin.com/content/dam/business/talent-solutions/regional/de-de/c/pdfs/BigDataimPersonalmanagement_LinkedIn_Bitkom.pdf [2017-10-17].

94 Cf. Grimm/Maiß 2015.

95 For example, the manufacturer of the fitness bracelet "Fitbit" (see https://www.fitbit.com/de/product/corporate-solutions [2017-10-17]) or the start-up Soma Analytics, which offers apps that measure users' stress levels and then pass the data on to HR management (see http://www.soma-analytics.com [2017-10-17]).

health programmes. Fellow policyholders also benefit from the financial resources that are saved, which can be used for other services, and co-workers are relieved of the need to fill in for others who are sick.[96] Nonetheless, the risks associated with the collection and usability of large volumes of health-related data in this context must not be overlooked. Premium adjustments or warnings made for unhealthy behaviour are not in the interest of the data provider. Even data collected anonymously can be linked to certain individuals with relative ease, if it is combined with additional information such as, for example, the data obtained by an analytics firm from a user's online digital footprint. Thus, in the USA, Walmart and Time Warner have already begun using the services of health analytics firms like Castlight Health in order to garner information about possible pregnancies amongst their staff.[97]

2.5.4 Commercial use of health-related data by global IT and internet companies

A major proportion of data collection and use is in the hands of large, global companies such as Facebook, Google (Alphabet), Apple, Amazon or Microsoft. These corporations act primarily as service providers, offering – on the basis of their access to enormous volumes of data and the necessary infrastructure – search engines, interactive information platforms, services such as online shopping and a wide selection of multifunctional devices. They are financed either primarily or in part through the personalised advertisement of products directed at specific target groups and/or context-sensitive marketing. To this end, various kinds of user data are collected, stored and exploited in vast quantities. For such companies, which

96 In Great Britain such advantages are prominently advertised through the "Britain's Healthiest Workplace" initiative (see https://www.vitality.co.uk/business/healthiest-workplace [2017-10-17]).

97 Cf. http://fortune.com/2016/02/17/castlight-pregnancy-data [2017-10-17].

are also increasingly active in the healthcare sector[98], it is thus especially feasible to combine data of a primarily health-related nature with numerous other forms of information.

These companies offer software, hardware, technology development and online services for big-data applications and provide data-oriented businesses and other institutions systems, algorithms, devices and infrastructure for data collection, analysis, management and storage, with which processes can be accelerated and improved in order to ensure a highly efficient use of any relevant information. Even if such companies have thus far not specialised in health-related applications, they nonetheless play an important role in the area of big data and health. Amazon Web Services, for example, as the world's leading provider of cloud computing services, has contracts with hospitals, universities and pharmaceutical companies[99], and Google works with Mayo Clinic in the USA to improve search results on diseases and their symptoms.[100] Because the data generation is often user-driven, a positive feedback loop occurs (more users – better products – even more users), which also translates into a competitive advantage for large companies with many customers or users. To this can be added other scaling effects connected with the technological know-how and global activities of these services. Data, both in terms of its quality and its relevance, is thus an essential factor in the success of certain internet services and their providers.

Data-driven optimisation extends to all commercial functions of the internet, in particular the areas of e-commerce and online advertisement, on which the vast majority of internet business models are based. Because user data is systematically translated into monetary value by means of the "advertising

98 Examples are Google's health platform Google Fit, Google's sister companies Verily Life Sciences and Calico, Microsoft's health data storage HealthVault or Apple's ResearchKit and CareKit.

99 Cf. https://aws.amazon.com/de/health [2017-10-17].

100 See https://www.mayoclinic.org/giving-to-mayo-clinic/your-impact/features-stories/google-works-with-mayo-clinic-to-share-health-knowledge [2017-10-17].

optimisation machinery" operating in this domain, advertising markets have been coupled with data markets. Through various mechanisms of user identification (see section 2.4.1), data can be linked and enriched such that the best form of targeted advertisement can be more precisely deployed. A small number of services dominate here, one reason for which is the formation of data monopolies, which make it more difficult for competitors to gain a foothold in the market.

There are a number of health-related fields in the private sector where the subsidiaries or close partners of large IT corporations are active. Numerous firms in the health, fitness and lifestyle sector work with personal data in order to provide a wide spectrum of individualised services designed to encourage or assist in the structuring of healthy everyday routines. This is done both in cooperation with institutionalised medicine, as is the case with manufacturers of medical products and devices, as well as independently, as is the case with products aimed at personal health-management.[101] In exchange for their data, and with technical assistance in the form of wearable devices, electronic diaries and apps, users are to be empowered to realise their individual lifestyle preferences, to attain health and fitness goals and to network, exchange and compare themselves with others. Furthermore, at least in other countries, private companies are already making efforts to access health-relevant data belonging to public institutions, e.g. via cooperation agreements.[102]

---

101 Through Google Analytics and other services, Google has a wealth of health data that is shared with pharmaceutical companies to improve medical devices. For example, Google is working with the biotechnology company Biogen to improve therapies for multiple sclerosis and Alzheimer's and is working with the Swiss company Novartis to develop a contact lens capable of measuring blood sugar (see https://www.medizintechnologie.de/infopool/politik-wirtschaft/2015/der-angriff-der-giganten [2017-10-17]). IT companies offer numerous apps and wearables for personal health management. For example, the Apple Watch can use sensors to analyse the wearer's heart rate, pulse, and blood pressure. Fitness trackers record users' steps, calories burnt and distances covered. Nutrition apps, such as eatSimply, count calories according to body measurements and compile individual nutrition plans.

102 Cf. McGoogan 2017.

The growing activity of digital firms in the healthcare sector could be regarded as beneficial for research, for major internet companies have access to significantly larger pools of data and are equipped with more powerful analytical capacities and better technical and financial resources compared to the public sector.[103] However, because some companies deny data providers full access to their own data, as well as the ability to transfer and use this data, there is reason to fear that these developments will in fact hinder medical progress and reduce fairness of opportunity and participation (see section 4.5), and for two reasons: firstly, because such forms of non-transparent and isolated data analysis are largely removed from critical reflection, and secondly, because the overwhelming market dominance of these companies would tend to thwart a methodological plurality in studying and interpreting health data.

2.5.5 Collection of health-related data by individuals

The success of data-driven businesses and applications, including in research, is crucially dependent on the generation and sometimes targeted collection of data by individuals. Citizens and patients record increasingly diverse personal and health-related data by means of sensors in mobile devices and store, manage and share this data online. Such mobile health applications (m-health applications) encompass a heterogeneous array of health and fitness apps covering health promotion, prevention, diagnostics and therapy management. In 2015, there were already over 380,000 apps available worldwide related to sport, lifestyle, diet, medicine, health, and fitness.[104] Alongside smartphones, ever-increasing numbers of portable devices, outfitted with sensors, are being marketed and used

103 Cf. Wilbanks/Topol 2016.
104 Lucht/Boeker/Kramer 2015, 6.

(so-called wearables). In 2017, an estimated 310 million such devices were sold worldwide.[105] This market is characterised by a highly dynamic and confusing array of manufacturers, services, products, data formats and terms and conditions of use, and is still largely unregulated.

On the one hand, health and fitness apps are specifically used in the context of a new trend towards self-tracking (the quantified self, see below); on the other hand, because of a growing general interest in topics such as diet and fitness, they are enjoying increasingly widespread application beyond more narrowly dedicated circles of users.[106] There are also apps and devices that, in the context of telemedicine, are tailored to the needs of specific groups, e.g. chronically ill patients or healthcare professionals. Numerous apps and portals designed to monitor and regulate the blood-sugar levels of diabetes patients, for example, already follow this approach.[107] Some of these apps feed their data into healthcare systems and databases in the form of, for example, vital statistics recorded by patient self-monitoring devices or data from a patient's digital diary; this data is then shared with therapists or insurers.

In light of the new possibilities created by big data, individuals – as citizens, patients, perhaps as test subjects, and certainly as users of devices and of the internet and as customers of companies active in the digital sector – see themselves confronted with a diverse array of roles and interactions with other actors. For them, the possibilities and potential consequences arising from the collection of their health-related data can be difficult to grasp in a number of respects, especially when it is far from discernible to what extent data is being analysed or whether it is, or could be made to be, relevant to health – e.g. when using internet search engines or social networks.

105 Cf. https://www.gartner.com/newsroom/id/3790965 [2017-11-27].
106 See, for example, the report by the Munich Digital Institute: Zhelyazkova et al. 2017.
107 Cf. http://diadigital.de/zertifizierte-apps [2017-10-17] and Kitsiou et al. 2017.

Many citizens regard the deliberate generation of individual health data in the context of sports, fitness or general health monitoring, e.g. by means of wearables, as beneficial to health and wellbeing. At the same time, it should be clear to many of these users that this data will not necessarily remain within the scope of personal control, especially if it is passed on to doctors, coaches, fellow athletes, and others. Today, it is in fact nearly impossible to avoid generating data that can be used by various interested parties for potentially economic purposes. If statistical surveys are reliable indicators, data protection and informational self-determination are widely valued across the population[108]; nonetheless, the digitisation of the world we live in is already so far advanced that everyday behaviours and forms of communication – even outside of social networks, lifestyle apps, and the like – entail the automatic production of data. At the same time, however, many value these products and services not only because they save time and money, but because they allow one to maintain important social connections despite spatial distance.

As mentioned above, a current trend contributing to the flood of health-related data is the production and use of sensors and apps capable of collecting, preparing and linking to existing databases ever-larger volumes of individual health-related data, as well as data pertaining to one's environment and daily activities. These sensors and apps can be integrated into mobile phones or smart watches, for example, or can be built into specially developed wearable devices or even everyday articles such as clothing, and can, depending on the equipment involved, collect physical data such as pulse frequency, temperature, or sleep or movement patterns. This data is normally then placed in high-volume storage (cloud) and processed so that it can later be referred to or used in analysis.

108 See for example Deutsches Institut für Vertrauen und Sicherheit im Internet 2017; European Commission 2017 and Mooy 2017.

These devices and programs can serve to provide those affected with access to their health information irrespective of time or place, or provide a basis for fact-based healthcare. They are also marketed as aiding the realisation of a modern, health-conscious lifestyle and for the promotion of personal well-being. Accompanied by slogans such as lifelogging, self-quantification or self-tracking, they sometimes are also expressions of a specific form of self-determination that is regarded by many as especially authentic. This includes the transfer of data to certain third parties. The existing healthcare system is seen as placing too little emphasis on individual responsibility and the active pursuit of healthy lifestyles, and doing too little to individualise medical care. Self-tracking is seen as a way to address both these perceived shortcomings.

While the self-monitoring of health-related parameters can indeed contribute to individual health and well-being and effect positive changes to one's lifestyle and self-image, the risks it entails must not be overlooked. Excessive self-control can contribute to an exaggerated drive for optimisation detrimental to personal health, as well as the medicalisation of "natural" life processes. Furthermore, it is questionable whether the use of such technology actually is an expression of personal sovereignty (see section 4.3). Self-tracking instruments can draw their users into a state of inner compulsion, a phenomenon that could be characterised as "self-induced heteronomy".[109] A further danger is discrimination against those who do not wish to take part in such self-tracking.

Self-tracking tools can also be used for scientific purposes, providing an important quantitative and qualitative expansion, and new levels of detail, to the data on which medical and health research is based. Apple's ResearchKit, for example, already contributes to epilepsy research.[110] At the same time,

109 Lob-Hüdepohl 2007, 84.

110 Cf. https://www.apple.com/de/newsroom/2015/10/15Apple-Announces-New-ResearchKit-Studies-for-Autism-Epilepsy-Melanoma [2017-10-17].

the providers and operators of such devices and apps as well as other companies have a major economic interest in the collection of such data, in order to make use of it by processing, linking and generating new correlations for others, to their own ends, whatever these may be. The fact that these services remain chiefly oriented towards this economic interest – alongside a deficit of user-friendliness, transparency and data protection – has been criticised, and a stronger orientation towards the needs of users has been demanded.[111] Accordingly, technical and other measures must be taken to strengthen transparency and control, reduce unrestricted access to data, clarify the scope of data use, and make informed consent possible and accessible.

2.6 Interim conclusion

The above discussion has shown that the use of big data in clinical practice, medical research and in the wider health-related domain entails certain strengths and weaknesses as well as opportunities and risks. These can be outlined as follows, using a SWOT analysis approach:[112]

111 Cf., for example, Albrecht 2016.

112 SWOT analysis is a situation analysis instrument that is primarily used in marketing, in management, in personality development, but also in political strategy development and strategic analysis in science management. The analysis focuses on comparing and weighing the strengths, weaknesses, opportunities and threats pertaining to a given or planned situation (cf. Pelz 2017 and Schröder 2011, 43-50).

Strengths	Weaknesses
1. Expansion and diversification of usable data as well as acceleration of information gathering 2. A positive feedback between the development of innovative data processing tools and the expansion of databases 3. A high degree of networking and ubiquitous accessibility	1. Heterogeneous data quality 2. Non-transparency of data flows and loss of control over data 3. Greater demands in terms of coordination, regulation and qualification
Opportunities	**Risks**
1. Refined stratification in diagnostics, therapy and prevention underpinned by a broadened knowledge base 2. Increases in efficiency and efficacy 3. Support for behaviour that promotes health	1. Solidarity erosion and diffusion of responsibility 2. Monopolisation and data misuse 3. Informational self-endangerment

The above reflections have however also revealed that any concrete judgment of big-data applications in connection with health and the healthcare sector must depend significantly on the actors involved, each with their own interests and assessments of risks and opportunities, as well as the wider societal context. Furthermore, it is necessary to identify who benefits from and who is negatively impacted by data collection and use, what the benefits and/or costs consist of, whether the benefits can be regarded as significant or the costs bearable, and what the likelihood is of certain scenarios.

The following legal (see chapter 3) and ethical (see chapter 4) analyses take into account the general opportunities and risks outlined above as well as context-specific factors. The responsible shaping of informational freedom on the part of the individual will then be developed as the key normative standard according to which these issues can be afforded balanced consideration (see chapter 5).

3 LEGAL GUIDELINES FOR BIG DATA

The myriad of at least potentially health-related data collected in various and often overlapping fields and contexts, and the many ways in which this data can be used, pose a considerable challenge to the legal system. Certainly in the healthcare sector, the opportunities and hopes associated with big data, outlined in chapter 2, are clearly recognisable. At the same time, however, there is a legitimate distrust in the quality, reliability and validity of health applications based on big data. The principle of counteracting information asymmetries, which underlies existing (data protection) law is thus by no means obsolete in the face of big data. On the contrary, "The increasing amount of data becoming available and processed as an effect of the tendency to rely on Big Data will only magnify the information asymmetry and increase the divide between controllers and users."[113] What is needed is a regulatory framework that will ensure the adequate safety and reliability of new and rapidly evolving technological applications, while at the same time providing them and the opportunities they present with sufficient scope to be developed.

When analysing and taking stock of the relevant legal controls, an important distinction must be made between constitutional and non-constitutional directives. The former also have binding consequences on the legislative and can thus restrict its scope of design in creating new laws. Nonetheless, both German and European constitutional law define a fairly open regulatory framework. As a rule, one cannot interpretively derive from this framework any detailed problem-solving models. It would thus be a fallacy to interpret currently applicable data protection law as the most authentic concretisation of constitutional norms, for which there exists "no alternative" (see section 3.2). Precisely because there exists considerable

113 European Data Protection Supervisor 2015, 9.

room for manoeuvre, it is necessary to further investigate the existing non-constitutional regulatory regime and its (dys) functionality with regard to big-data uses (see section 3.2). Thus, it is appropriate to concentrate on general data protection law, special data protection regulations pertaining to the health sector and the medical device law. The underlying incentive mechanisms must also be analysed. In particular, it must be determined whether and to what extent services based on big data fit into the system of (private and statutory) health insurance.

This analysis, which concerns itself with functionality and/or deficits in existing legislation, must then be supplemented with an overview of the regulatory concepts that may be sensibly applied in this area in the future (see section 3.3). As a result of the issue of territorial scope, which is an obvious problem for data-related, digital services, but also in view of the extreme dynamism of this sector, alternatives to the classic models of sovereign action must be explored. Rather, private (self-regulatory) as well as hybrid control mechanisms deserve more attention. However, at this point the emphasis is on more abstract observations and basic technical statements about the different legal instruments under consideration. Individual policy proposals are included in the recommendation section, taking into account the reflections of the legal and ethics chapters.

3.1 Control directives grounded in basic rights

The essential components of data protection law are grounded in basic, constitutional rights. The core constitutional standard at the national level is the so-called right to informational self-determination, a principle elaborated by the Federal Constitutional Court, in a landmark 1983 judgement on census participation, as a specific instantiation of the general right of

personality, in accordance with Article 2 (1) in conjunction with Article 1 (1) GG (Basic Law).[114] At the level of the European Union, the provisions of Articles 7 and 8 of the EU Charter of Fundamental Rights, as well as Article 16 (1) of the Treaty on the Functioning of the European Union, assert their constitutionally normative, directive power in a largely parallel manner. The same applies to the data protection principles developed on the basis of Article 8 of the European Convention on Human Rights.[115]

Already in its 1983 census ruling, the Federal Constitutional Court, with a keen sense of the consequences of technological progress, emphasised that due to the "data processing and linking capacities specific to information technology" there "no longer can be said to exist an 'irrelevant' item of data".[116] On the other hand, the judgement also asserted that a social and legal order "in which citizens can no longer know who knows what about them and when," is irreconcilable with the right of informational self-determination, a position that is undoubtedly excessive in light of the possibilities for decontextualisation and recontextualisation opened up by big data.[117] Contrary to this misconceived formulation, by which the right to informational self-determination takes on the appearance of a property, this right actually has an instrumental character. It serves the right to freely develop one's own person in accordance with the double guarantee contained in Article 2 (1) GG. The right to informational self-determination is not only aimed at the protection of general personality rights[118], but also at ensuring general freedom of action and other components of Article 2 (1) GG.[119] The Federal Constitutional Court has also repeatedly emphasised in recent decisions that the right to

114 BVerfGE 65, 1.
115 In the context of health data protection law, see Kühling/Seidel 2015, 155 ff., 162 ff.
116 BVerfGE 65, 1 (45).
117 BVerfGE 65, 1 (43).
118 According to the traditional reconstruction.
119 For example, (Federal Constitutional Court Justice) Britz 2010, 573 ff.

informational self-determination flanks and extends the constitutional protection of privacy and freedom of conduct.[120] Contrary to widespread misinterpretation, it must be emphasised that the right to informational self-determination also includes the power to decide with what content and through which relationships one wishes to enter into the process of personal development in interaction with one's environment. This applies to both German and European constitutional law.[121] These freedoms, in turn, may conflict with important public interest concerns such as the promotion of scientific progress or the provision of effective healthcare. However, the fundamental rights of other individuals, subject to private law, who wish to access and process available information must also be taken into account as potentially opposing considerations. In view of these potential conflicts, constitutional provisions merely form a regulatory framework, the details of which are left to the legislative.

As far as sovereign restrictions of the right to informational self-determination (including privacy) are concerned, the state must observe the negative function of its citizens' fundamental rights, while respecting the principle of proportionality. In this respect, constitutionally governed data protection law thus not only asserts the right to intervene legally in citizens' fundamental rights, but also contains more "modern" elements such as institutional safeguards (e.g. data protection officers) and procedural and transparency requirements.

However, big-data problems often involve horizontal legal relationships between private entities endowed with fundamental rights.[122] Data protection is not an end in itself, but serves the informational self-determination of the persons

120 For example, BVerfGE 120, 378 (397).

121 With regard to Article 8 (1) CFR, for example, Clement 2017, 169; cf. also BVerfGE 120, 180 (197): The general right of personality also has as its aim "the maintenance of the basic conditions of social relations between the holder of the fundamental right and his environment."

122 See also Buchner 2006, 46 ff.

involved. Accordingly, the state has a fundamental duty to take action against private threats to privacy.[123] Data protection law responds to this need, for example, by nullifying the force of consent when data is processed with a lack of transparency and accountability.[124] The Federal Constitutional Court has called for obligatory cooperation between users and data processors to determine in dialogue which data is necessary in a given instance.[125] However, the manifold and constantly changing forms of informational self-endangerment[126] that exist and the subtle ways in which power is exercised in data relationships must also be more closely taken into account (see section 4.3). These issues demand from lawmakers, whose remit is to balance the freedoms at stake, approaches that are suited to the complexity and dynamism of this regulatory field. Here again it is evident that constitutional law cannot provide immediate or conclusive answers.

3.2 Provisions in non-constitutional law

Because of the novelty of the phenomenon, at present only general legal provisions exist for health services based on big data. Special norms that explicitly address and act to structure the fundamental shift towards big data are almost completely absent. Legal analysis must therefore begin with the question as to whether and how novel technical possibilities fit with existing regulations that were originally created for other contexts. It must be determined whether and to what extent applicable law is prepared to appropriately address the new constellation of issues surrounding big data and where tensions arise. For

123 See, for example, BVerfGE 84, 192 (194 f.); also, for example, Rudolf 2011, para. 26 ff.; for the performance-related content of Article 8 (1) CFR, Article 16 (1) TFEU, see also Augsberg 2015, para. 8.
124 See, for example, Simitis 2014a, para. 7.
125 See BVerfGE in NJW 2013, 3086 (3088 f.).
126 See on this Hermstrüwer 2016.

this purpose, general data protection law and data protection law in the healthcare sector will be examined more closely, followed by a consideration of medical device law and health insurance law.

3.2.1 Big data as a challenge for applicable data protection law

As the following discussion will show, even after recent revisions occasioned by the enactment of the European General Data Protection Regulation (GDPR)[127], data protection law remains insufficiently adapted to the phenomenon of big data. Undoubtedly, current regulatory efforts have yielded significant improvements in the area of data protection. The GDPR and the so-called E-Privacy Regulation, which is still within the legislative process at the European level[128], represent clear progress, particularly with regard to the establishment of cross-border standards and the greater involvement of the concept of privacy by design (Article 25 GDPR)[129]. Despite criticism voiced in this regard, this is in principle also true of the German legislation implementing the GDPR[130] and the new

127 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (OJ 2016 L 119/1, 1-88).

128 See the Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications) of 10 January 2017 (COM[2017] 10 final). Following the recent majority vote of the European Parliament in favour of the (slightly amended) draft, negotiations are now taking place between Parliament, the Council of Ministers and the Commission, within the framework of the so-called trilogue process. For a thorough account, see Engeler/Felber 2017; for a critical viewpoint, see Hanfeld 2017.

129 For more details see Baumgartner/Gausling 2017; Gossen/Schramm 2017 and Schmitz/Dall'Armi 2017.

130 Gesetz zur Anpassung des Datenschutzrechts an die Verordnung (EU) 2016/679 und zur Umsetzung der Richtlinie (EU) 2016/680 of 30 June 2017 (BGBl. I, 2097).

version of the Bundesdatenschutzgesetz (Federal Data Protection Act, BDSG) it contains. Nonetheless, it does not suffice to demand an adequate implementation of EU requirements among member states in a way that ensures legal interoperability.[131] Above all, it cannot be assumed that a "full implementation and enforcement of the data protection regulations applicable in this country" would also ensure "adequate data protection" in the face of modern big-data applications (here: health apps).[132] On the contrary, the fundamental assumptions, core principles and objectives of traditional data protection law can hardly be reconciled with the distinct 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. Big data thus cannot be smoothly adapted to existing legal provisions and the arrangements by which they are rendered concrete. This has significant consequences. 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.

Personal data

As mentioned above, in its 1983 census ruling the Federal Constitutional Court already emphasised that, "Under the conditions of automatic data processing, there can no longer be said

131 According to the European Commission (cf. Habl et al. 2016, 54).

132 Albrecht 2016, 29: "The weaknesses are likely to be in the implementation by the providers and the lack of transparency in clarification and obtaining consent, as well as the sensitivity of users towards data protection issues." The inadequate regulation of international app services and the lack of implementation of existing requirements are described as especially problematic.

to exist an 'irrelevant' item of data."[133] However, applicable data protection law orients itself toward the personal nature of data and the concomitant obligation to adhere to certain acceptable uses of this data. This applies to the GDPR[134], the BDSG (in both its old and new versions) and to protections of personal data under civil law that derive from the general right to the protection of personality.[135]

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. The volumes of data used in big data do not simply derive from a given, specific context of use, nor are they restricted to such a context. "A specific investigative approach with a fixed, predetermined set of questions is not designed from the outset; rather, data is collected without a given purpose, or at least without being bound to a given purpose, and is then scoured by various 'algorithms', which themselves are constantly evolving, i.e. designed to learn, for productive patterns, which in turn are supposed to yield statements especially regarding prognosis or suitable therapies."[136] The "open-ended nature of the possible linkages within big data" thus seems to call into question the very starting point of data protection law, namely, identifying and protecting data of a personal nature.[137]

Purpose Limitation

Furthermore, there exists a tension between big-data applications and the fundamental requirement of data protection law that the use of personal data be limited to a specific purpose,

133 BVerfGE 65, 1 (45).
134 Cf., for example, Hofmann/Johannes 2017 and Krügel 2017.
135 Cf., for example, Werkmeister/Brandt 2016, 234 f. on personal references; see also the judgment of the European Court of Justice in NJW 2016, 3579 (3580 ff.).
136 Ladeur 2016, 360.
137 Ibid., 363 with reference to Tene/Polonetsky 2012, 63; Schwartz/Solove 2011, 1814. See also Schneider 2017 on the conflict.

a requirement again emphasised in the GDPR.[138] Accordingly, personal data may only be stored, altered or used in connection with the purpose for which it was originally collected, a requirement that aims to establish transparency and enable a level of control over the legality of data processing[139]. Closely linked to this is the principle that the objective laid out in the licensing and consent process can only be achieved through the use of personal data (see Section 28 (1) sentence 1 no. 2 BDSG, old version). However, this also contradicts the inner logic of big data, which is characterised by an openness of outcome and constant processes of reshuffling and reorientation with regards to an end result. In the world of big data, data that was stored for one purpose is often used for another. Accordingly, it is necessary for big-data applications to collect and store data for purposes that have yet to be determined.[140] In the context of the statistically based and probability-oriented means of generating insights typical of big data, a strict purpose limitation of data is thus hardly feasible. The principle of necessity that is fundamental to traditional data protection law thus loses its meaning: "If data is analysed for any arbitrary purpose, it is already necessary for that purpose."[141]

Data parsimony

Furthermore, big data is obviously incompatible with the principle of data parsimony/data avoidance – standardised in Recital 156, Article 5 (1) (c), Article 25 (1) GDPR and Section 71 (1) BDSG, new version – which is also referred to as data minimisation in the language of the GDPR. According to this principal as little personal data as possible is to be collected, processed or used. A corresponding obligation is to delete data as soon as it is no longer needed (see, for example, Section 35 BDSG, new version). As thoroughly described in chapter 2

138 Cf. Schantz 2016, 1843 f.
139 Cf. Simitis 2014b, para. 111.
140 Cf. Roßnagel/Nebel 2015, 458.
141 Roßnagel/Nebel 2015, 458.

above, however, the potential of big data is unleashed most effectively when data is gathered and then linked to the greatest extent possible. Deleting data would mean that certain correlations or patterns would not be discovered. A blanket adherence to the principle of data parsimony would therefore largely preclude tapping into the potential of big data. Nonetheless, because the danger posed to the right to informational self-determination increases with the amount of stored data, effective alternative safeguards are needed.

The consent requirement

The incompatibilities between big data and the traditional approach to data protection can also be illustrated by the requirement to obtain consent – as stated, for example, in Article 6 (1) and Article 7 GDPR – which dominates present-day data protection law.[142] Subject to a separate legal permission issued in relation to its clearly defined purpose, data use is thus only allowed if those affected have granted their consent. Granting consent is therefore the most important means for affected parties to exercise their right to informational self-determination; it is intended to restrict and control the use of personal data according to the will of the legislator.[143] In principle, the GDPR adheres firmly to this approach. While it does call for a relaxation of the relatively strict written form requirements under German law[144] as well as allowing for the possibility of issuing consent for "one or more given purposes" (Article 6 (1) (a) GDPR), it actually represents no real departure from the concept employed until the present. For according to the definition of Article 4 no. 11 of the GDPR (corresponding to Section 46 no. 17 BDSG, new version), the consent "of the data

142 Cf. Tene/Polonetsky 2012, 67, who are generally sceptical about the efficiency of individual consent in view of "on the one hand, well-documented cognitive biases, and on the other hand the increasing complexity of the information ecosystem."

143 See also Simitis 2014a, para. 4 for consent in accordance with Section 4a BDSG, old version.

144 Cf. Krohm 2016 as well as Thüsing/Schmidt/Forst 2017.

subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her."[145]

Against this background, it must continue to be assumed for both national and EU law that consent is only effective if the person affected is able to ascertain the meaning and scope of the intended use of the data. For this reason, they must be informed about the goals of data processing, the identity of the processors and the scope of the processing (see recital 42 of the GDPR). If data is transferred, the consent granted must define to whom this data may be made accessible.[146] Furthermore, information about the data concerned, the intended duration of storage, and the consequences of withheld consent are required as the basis of adequate decision-making.[147] Particularly with a view to especially troublesome constellations of data use, Article 8 of the GDPR now contains an area-specific development of the existing consent model. To date, according to both EU and national data protection law, effective consent did not presuppose the full legal age of the person granting

145 See also Recital 32 of the GDPR: "Consent should be given by a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the data subject's agreement to the processing of personal data relating to him or her, such as by a written statement, including by electronic means, or an oral statement. [...] Consent should cover all processing activities carried out for the same purpose or purposes. When the processing has multiple purposes, consent should be given for all of them. If the data subject's consent is to be given following a request by electronic means, the request must be clear, concise and not unnecessarily disruptive to the use of the service for which it is provided." See also Buchner/Kühling 2017, para. 61 ff. However, more generous with regard to the area of "scientific research" is Recital 33. It states that, where the purpose of the processing cannot be fully stated at the time of collection of the personal data, "data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognised ethical standards for scientific research. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose."

146 For Section 4a BDSG, old version, see also Simitis 2014a, para. 80, and for Article 7 GDPR Buchner/Kühling 2017, para. 20 ff.; see also Schaar 2017.

147 Cf., for example, Sivridis/Seidel/Kühling 2015, para. 319.

it, but merely their sufficient cognitive faculty, to be determined on a case-by-case basis.[148] The revision of the Article 8 of the GDPR responds to the fact that minors are less aware of the risks associated with the processing of personal data than adults, mandating a blanket increase in protection. For the, practically speaking, especially relevant field of "information society services,"[149] the standard establishes a minimum age of consent of 16 years. If this age is not reached, the legal guardian must always be included in the granting of consent.[150] Article 8 (2) of the GDPR also obliges data users to adopt technical and organisational measures to ensure that "consent is given or authorised by the holder of parental responsibility over the child." However, an (age) verification system is not specified; in addition, in accordance with the principle of proportionality, this obligation is limited to "reasonable efforts."[151]

The functionality and effectiveness of these requirements remain open to debate. Not only among minors, but also among adults, it is often more than doubtful that data providers truly understand the purposes for which their data is used and the implications thereof. From a realistic perspective, consent is thus often referred to as a mere formality. Especially in the healthcare sector, the notion of a truly autonomous decision on the part of the patient is frequently called into question

148 Cf. Ernst 2017, 111 with further references and Kampert 2017, para. 2 with further references.

149 According to Article 4 no. 25 GDPR, services within the meaning of Article 1 no. 1b of Directive 2015/1535/EU are covered, i.e. (only) "any service normally provided for remuneration, at a distance, by electronic means and at the individual request of a recipient of services." It is important to note that remunerated services in this sense also include services financed by advertising, which are therefore free of charge for the user themselves (cf. Kampert 2017, para. 268 with further references). On the other hand, the explicit exception granted to prevention or counselling services that are offered directly to children (cf. Recital 38 of the GDPR), which is obviously in the best interests of the child, is likely to be primarily declaratory in character because these services are as a rule free of charge.

150 Article 8 (1) sentence 3 GDPR allows the Member States to derogate from this requirement up to an absolute lower limit of 13 years. However, the German legislative has not made use of this opening clause.

151 Cf. Kampert 2017, para. 13 f.

by the knowledge asymmetries that exist between patients and healthcare providers. This is particularly the case when the disclosure of data is made a prerequisite for the utilisation of certain medical services, thus deepening the existing relationship of dependence between medical institutions and affected persons.[152]

Big-data applications significantly compound this general problem.[153] First of all, the future uses to which data may be put are not even vaguely discernible at the time data is collected, making it impossible to inform those affected about how their data will be employed.[154] Big data not only enables real-time monitoring, but also the simulation and prediction of future scenarios.[155] Thus, those affected cannot foresee the results that may be generated by their data. Furthermore, temporal limits can hardly be placed on analyses using big data. Granting blanket consent for one's data to be used for any purpose, however, does not suffice under current data protection law, as it would amount to a merely abstract commitment to self-determination.[156]

Even assuming that the party or parties processing data would be able to at least partly predict the way this data will be used, it remains doubtful that such complex and extensive information could at all be communicated to data providers in a way that they could understand. At present, many data consent agreements are so detailed and opaque that the consumer is simply overwhelmed with information.[157] In practice, consent agreements of this kind do not provide a sufficient basis of legitimacy. In the context of big-data applications, the granting of consent thus degrades to the point of mere formality and is hardly suitable as a means of legitimation.

152 Cf. Augsberg 2016, on the example of the transplantation register.
153 Cf. Becker 2017, 173, who is thus critical; Ulbricht/Weber 2017 and Mostert et al. 2016.
154 Cf. Mayer-Schönberger/Cukier 2013, 171 ff.
155 Cf. Langkafel 2015, 27.
156 Cf. Ernst 2017, 113, and Simitis 2014a, para. 3, 77.
157 Cf. Katko/Babaei-Beigi 2014, 362.

Overall, it is clear that applicable data protection law is focused on consent and thus only acts on a case-by-case basis. Apart from the right of revocation (Article 7 (3) GDPR) and the rights to access, rectification, blocking and deletion (which will be considered below), it offers little scope for influencing the ultimate fate of one's data in a processual manner. In principle, the use of data for purposes other than those covered by an existing consent agreement always requires a new consent agreement, and this consent must be granted before data processing begins.[158] Subsequent declarations cannot reverse the inadmissibility of processing without prior consent, for authorisation (Section 184 BGB [Civil Code]) is not the same as prior consent (Section 183 BGB), and the BDSG does not provide for curing such a breach.[159] Once the data is collected with consent, it can no longer be followed by the person affected. Of course, the rights to access, rectification, blocking and deletion (e.g. Article 13 ff. GDPR, Section 32 ff. BDSG, new version)[160] serve those affected as rights of control, allowing them to effectively exercise their right to informational self-determination and granting them the ability to determine whether their data has been used in accordance with the intended purpose. Because data providers are entitled to the cessation of inadmissible data use and compensation for any damages incurred as a result, data misuse can thus be averted or sanctioned. However, these instruments are designed in such a way that assumes that the questions put to the data or the will of the data provider are static and delineated (e.g. relating to a given dataset or timeframe), and are thus conceived more as case-by-case interventions. The dynamic of big data does not fit into this regulatory concept. Especially if one holds the consent of those affected to be a basic, indispensable requirement of data protection, one

158 See Simitis 2014a, para. 27.
159 Cf. ibid., para. 29.
160 Cf. BVerfGE 65, 1 (46): "Further procedural safeguards include the duties to clarify, to provide information and to delete information."

must search for ways to make consent functional and meaningful even under the conditions of big data.

Anonymisation and pseudonymisation

In the face of the possibility of reidentification (facilitated by computers and data processing and thus intensified under the conditions of big data), there exists a fundamental doubt as to the effectiveness of the requirement of data anonymisation and pseudonymisation.[161] Granted, the requirement to eliminate wherever possible any connection data may have to a person does not in principle stand in contradiction to the concept of big data. On the contrary, even if data can be deemed to be of a decisively personal nature, based on the conventional understanding of whether data protection law applies to a specific situation, this is generally insignificant for big-data applications.

Nonetheless, in the case of medical data in particular, it is fundamentally necessary to ask whether the personal connections data may have can ever be completely avoided or eliminated in the age of big data. It has been argued in the past that anonymisation of genomic data is in fact impossible due to the unique "genetic fingerprint" of every human being.[162] Already, biological, clinical and socio-demographic data, as well as lifestyle data (e.g. one's degree of physical activity, eating habits, etc.), are collected by researchers investigating the causes of complex diseases.[163] In the future, as part of a systems-oriented approach in medicine, comprehensive molecular data (so-called omics data, see section 2.4.2), clinical data socio-demographic data, psychosocial data and data related to lifestyle and environmental influences should be collected in a therapeutic context for every patient, in order to better analyse and predict their state of health. Thus, patient datasets will

161 Cf. Ohm 2010 more generally; Mostert et al. 2016 more specifically to the area in question.
162 Cf., for example, Berdin 2017, 194; Vossenkuhl 2013, 97.
163 Kollek 2012, 26.

become increasingly individualised. In order to ensure effective anonymisation, various parameters would have to be stored independently before being transferred into another context. However, this contradicts the central concept of big data, i.e. that novel statements regarding the discovery of unknown correlations and connections can be made precisely because of the total interlinking of the most diverse kinds of individual data. Every time data is transferred across contexts, even when dealing with partially separated datasets that have each been anonymised, there is a risk that anonymised patient or volunteer data could be linked back to individual persons by means of analysis and combination with other data available online or contained in any number of other databases.[164] The more individual a dataset, the more easily it can be concretised and ultimately traced back to an individual person with the help of supplementary information contained elsewhere. It does not matter which data sources could "reasonably" be considered significant or whether the access to this data is legally permissible or inadmissible.[165] Especially risky in this context is the myriad data that consumers themselves create, both knowingly and unknowingly, through their online activity.[166] Overall, as a result of the improvement of analytics and the expansion of the pool of available data, the effort required to reidentify data with its provider is reduced every year.[167] The question regarding at which point and to what extent the danger of reidentification surrounding a given item of anonymised data should be considered sufficient[168] only intensifies the problem

164 Cf. Weichert 2014a, 170.
165 Cf. ibid.
166 Cf. Chatziastros/Drepper/Semler 2014.
167 Cf. Weichert 2014b, 836.
168 With regard to anonymisation, data protection law focuses on whether the data can be assigned to a specific or identifiable person "only with a disproportionately large expenditure of time, cost and labour" (cf. Section 3 (6) BDSG, old version).

of defining what constitutes personal data for the purposes of data protection law.[169]

Information, correction, deletion and blocking

The rights to be informed about the collection of personal data, as well as to its correction, deletion, and blocking, mentioned above (Article 13 ff. GDPR, Section 32 ff. BDSG, new version) apply insofar and as long as data exhibit a personal connection.[170] They therefore serve the principle of transparency. However, they often do not provide effective protection because they are virtually impossible or at least very difficult to enforce, and are thus de facto idle.[171] Even defining the scope and depth of data regarding which one would have the right to be informed is a problematic task, in light of the multitude of data points that those affected cannot grasp in their totality and from which new insights are constantly being generated. As with consent, the right to information also in practice entails a considerable (and occasionally problematic) need for additional information and possibly even medical advice on the part of those exercising it.[172] Furthermore, a data provider can hardly know all potential parties with competing claims, which would be necessary in order to gain a comprehensive grasp of the total, scattered mosaic of information about them in circulation at any given time. For precisely this reason, there is justifiable reason to doubt that the right to data portability contained in Article 20 of the GDPR, and the hopes that have been pinned to it, can be truly realised. This provision grants to affected persons the right, among others, "to receive the

169 For example, the data protection authorities consider it sufficient for the assumption of a personal reference that any third party can reassign (pseudonymised) data to its provider. Thus, data would be personal for each recipient. The opposite view, on the other hand, focuses on whether the party storing data can establish a personal reference on its own. Thus, according to this position, data is only personal data for those who can establish the personal reference.

170 Cf. Weichert 2013, 252; see also Paal/Hennemann 2017, 1700 f.

171 Cf. also Keppeler/Berning 2017.

172 Cf. Weichert 2014b, 837.

personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format." This should allow "data subjects" better control over their own data while also ensuring that lock-in effects can be avoided and services easily changed.[173] However, it is precisely the act of establishing who the data "controller" is, i.e. identifying the party responsible for the data, that will prove most difficult under the conditions of big data.

A fundamental aspect of the right to be informed about one's data is the traceability of the data use process. In this respect, the provisions of Article 22 of the GDPR are particularly relevant. Because the individual should not be made the object of a purely mechanical decision, fully automated processing methods are expressly prohibited. The exceptions to this prohibition (involving contractual relationships, subject to special regulations and with the data provider's consent) additionally require suitable protective measures – in particular a ban on discrimination – and are tied to certain disclosure requirements.[174] Affected persons must be able to discern which data and analytical methods are involved in processing and how the output of this process has been produced.[175] This would apply in particular to algorithms. However, algorithms themselves are not subject to the right to information, due to the intellectual services they provide and for the protection of business secrets.[176] Even if the algorithms being employed in the healthcare sector were to be made transparent for the purposes of comparison, because of the barely comprehensible, complex nature of the formulas they contain – especially in the case of self-training algorithms in the field of deep learning (see section 2.3.2) – the individual's right to information would

173 Cf. Keßler 2017, 591 and Paal/Hennemann 2017, 1701.
174 Cf. Buchner 2017, para. 11 ff. and Martini/Nink 2017.
175 Cf. Martini 2014, 1484.
176 Cf., for example, Helfrich 2017, para. 77 ff. on the right to information pursuant to Article 22 (1), Article 15 (1) (h) GDPR; see also Martini 2014, 1485; also the judgment of the Federal Court of Justice in NJW 2014, 1235 ff.

hardly be served. Because algorithms can also produce erroneous conclusions, this lack of transparency is highly problematic in terms of data protection law, especially as there exists no state control of algorithms.[177] As a result, the existing right to rectification and erasure is revealed to be idle, as the person affected cannot exercise these rights without possessing comprehensive information.

3.2.2 Health data protection law

The deficits analysed above in relation to general data protection law can be said, with only some restrictions, to apply equally to data protection laws pertaining specifically to health data. This area is characterised by a double security regime: on the one hand, personal data in the healthcare sector is protected by, in part, specially designed provisions of data protection law; on the other hand, as patient data, it is subject to the provisions of civil, criminal and professional legislation pertaining to the obligation to medical confidentiality.

Health data protection law is, however, bound to the same basic logic as general data protection law: it is similarly focussed on consent and relies upon comparable regulatory approaches (compare, for example, the pertinent regulations contained in the GenDG). Of course, the consent model has been modified, e.g. in cancer registry regulations[178], and biobank regulations rely, in a legally questionable manner, upon the concept of global consent.[179] However, the normative solu-

177 Cf. Martini 2014, 1485.
178 For example, the disclosure of data for a specific research project without the consent of the person concerned on the basis of an official authorisation; see, for example, Section 9 Hamburgisches Krebsregistergesetz (Hamburg Cancer Registry Act).
179 See Torbohm 2015, 367.

tions to the problems of health data protection are for the most part still bound to a "pre-big-data" perspective.[180]

This is particularly evident in the traditional distinction between data of different sensitivity levels.[181] Conventionally, personal data related to health has been classified as especially sensitive and thus requiring special protection. Both the GDPR (Article 4 nos. 13, 15, Article 9, Recital 35, 52 ff.) and the federal and state data protection laws (see, e.g. Section 3 (9) BDSG, old version, Section 46 no. 13 BDSG, new version; see also Section 35 SGB I, Section 67 ff. SGB X and Section 284 SGB V) contain corresponding, separate provisions concerning health-related data.

The underlying assumption – namely, that the sensitivity of this data is established on the basis of its connection to health – does not apply under the conditions of big data, or at least only partially applies. Characteristic of big data are the above-mentioned possibilities of continuous decontextualisation and recontextualisation, unfolding either "in pursuit of new objectives (bottom-up) or through spontaneously generated correlations and patterns, which unleash a dynamic of knowledge production that proceeds by means of variable constellations, but do not depict 'reality'."[182] Given the possible combination of molecular and clinical patient data with psychosocial, socio-demographic, lifestyle and other data to produce a body of "total health information" for each patient, it is unclear whether these sorts of categorical distinctions can continue to be upheld. Furthermore, data that is prima facie not directly relevant to one's health, such as supermarket shopping habits, can acquire such relevance in the hands of big data. Big-data analysis operates almost exclusively on the basis of

180 On laws governing biobanks cf. Berdin 2017, 54 ff., 159, 379: The legal regulatory framework continues to be based centrally on the requirement of individual consent, so that no appropriate regulation of biobank research can be achieved.

181 Cf. Ohm 2015.

182 Ladeur 2016, 361.

statistical correlations and probability values. Thus, for example, information regarding someone's genetic predispositions could acquire a significant predictive power in combination with other, prima facie non-health-related (environmental) information and allow special risk profiles to be developed. Neither existing data protection law nor the Genetic Diagnosis Act are adequately tailored to the peculiarities of the analytical methods applied to molecular data in the field of human medicine or the capacities for linking data opened up by big data. Generally speaking, there exists a significant incongruity between the characteristics of big data and the existing normative requirements for the use of data. The latter are oriented – not only, but also in the healthcare sector – towards the use of limited amounts of data by legitimised users for the care of a given patient.

Another source of difficulty is the fact that the legal provisions for the protection of personal data in the healthcare sector are especially complex and difficult to understand. On the one hand, personal data in the healthcare sector, as patient data, is subject to the provisions of civil, criminal and professional legislation pertaining to the obligation to medical confidentiality. On the other hand, the specific, relevant provisions of data protection law are constituted by a complex combination of federal and state regulations. Thus, for example, private health insurance is subject (only) to general data protection law. With regards to statutory health insurance, by contrast, specific regulations contained in the Social Code (SGB V and X) as well as special state provisions (e.g. North Rhine-Westphalia's Gesundheitsdatenschutzgesetz [Health Data Protection Law]) apply, although it remains unclear whether and to what extent they also apply to healthcare service providers (e.g. practising physicians and medical care centres). For hospitals, the relevant provisions of state hospital law are also partly responsible for standardising special data protection requirements. If these don't apply, privately owned facilities are in principle subject to the BDSG, while public clinics are subject

to state data protection statues, which themselves partly refer to the BDSG. For facilities owned by the church, there exist data protection regimes specific to the church.[183] Overarching this complex matrix of laws are the provisions of the European GDPR, which take overall precedence.

3.2.3 Interim conclusion

There is thus a significant discrepancy between the requirements of traditional data protection law and the effective conditions of big data. Even with an extremely high level of regulatory and organisational effort, it would hardly be possible to effectively and smoothly apply these traditional regulatory mechanisms to the new situation presented by big data. This applies especially to the self-governing healthcare system, which is characterised by a plurality of different service providers that have partially conflicting interests and that undertake corresponding efforts at exerting influence (see chapter 1). However, if one does not wish to draw from this regulatory deficit the conclusion that the use of big data should be deemed fundamentally inadmissible in any case which fails to (fully) comply with the provisions of all applicable data protection law, alternative mechanisms which are suitable to this domain and which can guarantee an adequate level of protection must be sought. However, because a reduction of constitutionally stipulated minimum standards must be ruled out, these alternatives must be sought not only in the further development of data protection law, but also in the compensatory effects of other, domain-specific regulations.[184]

183 For a comprehensive account, cf. Karaalp 2016, 59 ff.

184 Cf., on the other hand, Paal/Hennemann 2017, 1698 f., and Drexl 2017 on general competition law.

3.2.4 Medical device law

Such a compensatory effect could initially be exerted by the provisions of medical device law, which fundamentally aims to regulate the free circulation of medical products while guaranteeing the safety, suitability and performance of medical devices for the protection of patients, operators and third parties. Because it concerns product safety, medical device law can also be regarded as belonging to consumer protection law.[185] The legal basis of medical device law can be found in both European law and national law. Relevant provisions are contained above all in the German Medizinproduktegesetz (Medical Devices Act, MPG) and in corresponding references through which three European guidelines – for active implantable medical devices[186], for in vitro diagnostics[187] and for medical devices[188] – are to be implemented.[189] Additionally, two regulations were recently introduced at the European level concerning medical devices[190] and in vitro diagnostics[191]. Both entered into force on 25 May 2017 and will be mandatory after a transitional period of three or five years, respectively.

Unlike medications, medical products do not require government approval. However, according to Section 6 of the MPG, a medical device can only enter into circulation or

185 Cf. for this and the following, for example, Zirfas 2017, 106 ff.; Pannenbecker 2013, para. 250.

186 European Council Directive 90/385/EEC of 20 June 1990 on the approximation of the laws of the Member States relating to active implantable medical devices (OJ 1990 L 189/17).

187 Directive 98/79/EC of the European Parliament and of the Council, of 27 October 1998, on in vitro diagnostic medical devices (OJ 1998 L 331/1).

188 European Council Directive 93/42/EEC of 14 June 1993 concerning medical devices (OJ 1993 L 169/1).

189 Directives 90/385/EEC and 93/42/EEC were amended by Directive 2007/47/EC of 5 September 2007 (OJ 2007 L 247/21).

190 Regulation (EU) 2017/745 of the European Parliament and of the Council, of 5 April 2017, on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (OJ 2017 L 117/1).

191 Regulation (EU) 2017/746 of the European Parliament and of the Council, of 5 April 2017, on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU (OJ 2017 L 117/176).

operation if it is certified and given a CE label.[192] For this to happen, the device must comply with the basic requirements listed in Section 7 of the MPG, specifically in Annex I, concerning medical device guidelines. These include the need for a product-specific risk assessment, the implementation of risk-minimisation procedures and a risk–benefit analysis. Additionally and most importantly, a conformity assessment procedure tailored to the risks of a given device must be successfully completed. This means in practice that, according to the risk category in which the device has been placed under Section 13 of the MPG[193], the manufacturer can either itself carry out this assessment procedure and establish the conformity of the device (risk class I), or it must involve a "notified body".[194]

Consequently, medical devices are not subject to any preventive state control, but merely to a form of downstream control involving market oversight at the hands of state authorities and a medical device monitoring and reporting system overseen by a higher federal authority (see Section 26 ff. MPG), namely the Bundesinstitut für Arzneimittel und Medizinprodukte (Federal Institute for Drugs and Medical Devices) and the Paul Ehrlich Institute (Federal Institute for Vaccines and Biomedicines, translator's note). Nonetheless, the MPG threatens sanctions (Sections 40, 41 MPG) for violations of the above-mentioned standards (Sections 4, 6 MPG), including, under Section 4, for negligent conduct.

With regard to big-data applications, it is of considerable importance that software can also be classified as a medical device. This requires an intended medical purpose as listed in

192 The CE marking is a mandatory conformity marking that indicates that a product complies with the harmonisation regulations of the European Union. By attaching this label to a product, the manufacturer declares to the authorities that the product complies with the regulations and has undergone the prescribed conformity assessment procedures.

193 This requirement does not apply to in vitro diagnostic medical devices.

194 For detailed information on the risk classification of stand-alone software, see Oen 2009, 57.

Section 3 no. 1 of the MPG.[195] Section 3 no. 10 of the MPG defines intended purpose as the use for which the device is intended according to the information provided by its manufacturer (Section 3 no. 15 MPG) in its labelling, operating instructions or advertisement material. From this regulation it follows that determining a device's intended purpose, and thus its qualification as a medical device, depends to a large degree on the information its manufacturer provides.[196] Only to the extent that the intended purpose given by the manufacturer reveals itself to be scientifically untenable or contradictory does the determination of the objective suitability of a device become a matter for the courts.[197]

Clearly, this can lead to considerable difficulty in distinguishing between medical uses and mere lifestyle or fitness apps: depending on the context in question, diet or exercise tips, for example, can also serve a medical purpose.[198] Furthermore, manufacturers can exclude from their products' intended uses therapeutic or diagnostic functions in the medical context. As interpretive guides, one can refer to the (legally non-binding) European Commission guidelines on standalone software[199] as well as the Commission's handbook on borderline and classification of medical devices.[200] Apps that do not carry out analysis, but which limit themselves to the mere collection of data and its graphical representation, serve no diagnostic or therapeutic purposes.[201] Thus, information systems that serve only to store, playback or transfer data also

195 "(a) the detection, prevention, monitoring, treatment or alleviation of disease; (b) the detection, monitoring, treatment, alleviation or compensation of injury or disability; (c) the examination, replacement or modification of the anatomical structure or physiological process; or (d) the regulation of conception." Cf. for example Heimhalt/Rehmann 2014, 200 ff. for more details.
196 Cf. Heimhalt/Rehmann 2014, 201.
197 Cf. the judgment of the Federal Court of Justice in NJW-RR 2014, 46 (47).
198 Cf. Heimhalt/Rehmann 2014, 202.
199 Cf. European Commission 2012b.
200 Cf. European Commission 2015.
201 Cf. Heimhalt/Rehmann 2014, 202 with further references.

do not fall under the MPG.[202] However, these systems become medical devices as soon as they begin storing comprehensive patient records for the purpose of optimising treatment.[203] Similarly, any software that merges and analyses patient data in order to provide a diagnosis, suggest a certain therapy, calculate a dosage of medication or calibrate laboratory values with reference values falls under the medical device law.[204]

However, the necessary criteria for clinical evaluation are not always clear in relation to m-health software. As far as the requirements for evaluation are concerned in detail, Section 19 (1) of the MPG refers to relevant guidelines and, where applicable, standardised norms.[205] Nonetheless, neither the MPG nor the current EU Medical Device Regulation contain provisions that fully eliminate the difficulties in distinguishing between medical and non-medical devices outlined above.[206] The new regulations primarily contain adjustments of the existing legal situation, but no changes to the basic regulatory approach. Thus, some medical devices are placed in higher risk categories and "notified bodies" and oversight authorities are subject to more stringent demands. The requirements concerning clinical evaluation are also more detailed and more comprehensively designed. Particularly interesting is Recital 19 of the EU Medical Device Regulation.[207] Accordingly, "[i]t is necessary to clarify that software in its own right, when specifically intended by the manufacturer to be used for one or more of the medical purposes set out in the definition of a medical device, qualifies as a medical device, while software for

202 Cf. Rübsamen 2015, 487; European Commission 2012b, 20.
203 Cf. Dierks, after Gärtner 2010, 17.
204 Cf. Rübsamen 2015, 487; European Commission 2012b, 11, 20.
205 "If medical devices comply with harmonised standards or equivalent monographs of the European Pharmacopoeia, or common technical specifications relating to the respective medical device, the presumption is that the provisions of [the MPG] are being met." (Section 8 (1) MPG).
206 Cf. on the draft, Heimhalt/Rehmann 2014, 203; Gassner 2015, 77.
207 Regulation (EU) 2017/745 of the European Parliament and of the Council, of 5 April 2017, on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (OJ 2017 L 117/1).

general purposes, even when used in a healthcare setting, or software intended for lifestyle and well-being purposes is not a medical device. The qualification of software, either as a device or an accessory, is independent of the software's location or the type of interconnection between the software and a device." This second sentence in particular makes clear that there now exists a clear awareness of the need to make a distinction between health apps that qualify as medical devices and lifestyle apps. When qualifying health-related apps as medical devices in the future, due regard will likely be required for this kind of deliberate negative delimitation.[208]

3.2.5 Big data services in the context of (statutory) health insurance

The provisions of health insurance law are also relevant here, not only in the sense of compensating for inadequate protection standards, but also in terms of ensuring that the opportunities associated with big data can be taken advantage of in a sustainable way. At present, this has less to do with apps that are offered by private or statutory health insurance providers.[209] Rather, from a legal point of view, three particularly sensitive issues should be highlighted, which will be examined below as exemplary of the system of statutory health insurance: first, the necessary adaptation of applications based on big data into the existing payment model; second, the compatibility of the knowledge produced by big-data analysis with the epistemic standards and rules governing benefits; and third, the specific possibilities and limitations of big data with regard to premium adjustments.

The first issue illustrates the extent to which compensatory (data) protection can also be secured by means of financing

208 According to Gassner 2016, 111.
209 On this topic, an overview is provided by Aumann/Frank/Pramann 2016.

arrangements that in themselves have little to do with data protection. It is well known that one of the most problematic aspects of the internet economy is the fact that many services are offered free of charge, while in fact treating user data as a financially relevant form of compensation. Against this background, it is clear that a separate form of payment, namely the inclusion of m-health applications in the payment schedules of both private and statutory health insurance providers, would at least help mitigate this problem; a secure source of financing would reduce the incentive on the part of those offering such services to exploit the data they collect in other ways, including disseminating it in an uncontrolled manner contrary to the spirit of data protection law. That this is possible within the current legal framework is shown by pilot projects in the field of so-called telemedicine, such as outpatient ophthalmological care for children with disturbed development of binocular vision, for whom web-based stimulation therapy, including apps for tablets and smartphones, has been introduced. For the time being, this service is paid for by health insurers, not in the framework of standard care, but on the basis of specific contractual provisions (in the sense of Section 73c SGB V, which has nonetheless since been repealed) between insurers and a qualified network of ophthalmologists.[210]

There are no general objections to the inclusion of services based on big data in the standard coverage offered by statutory health insurance, as long as the basic principles of public health insurance are observed. This applies with greatest force to the solidarity principle, a cornerstone of the statutory health insurance system, which demands a certain "blindness" towards some kinds of knowledge (e.g. regarding pre-existing conditions or genetic dispositions). By definition, private health insurance does not adhere to such a strict principle of solidarity. However, or rather precisely for this reason, special care must be taken in the realm of private insurance to prevent

210 See GKV-Spitzenverband 2014, 11.

admissible forms of differentiation to slip into inadmissible forms of discrimination or stigmatisation.

The overarching need to conservatively utilise the limited resources of the healthcare system imposes additional requirements. The complex programme by which the legal parameters in this domain are put into practice deliberately encompasses the findings of medical science. Accordingly, Section 2 (1) sentence 3 SGB V fundamentally demands that the "quality and effectiveness of benefits […] must correspond with the generally accepted state of medical knowledge and take account of progress in medical science". With this reference, the legal system acknowledges the limits of its own knowledge and correctly understands science as a dynamic, open-ended process. However, no precise requirements can be extracted from this law with regard to how this scientific imperative can be fulfilled.[211] In practice, different strategies can be observed: for the time being, rulings in social courts tend to restrict themselves in part to deferring to expert advice in the classic sense, especially by searching for a consensus or at least a majority opinion among those with relevant expertise. However, the partly explicit (e.g. in Section 35 (1) (b) sentence 4, Section 35a (1) sentence 7 no. 2, Section 35b (1) sentence 5, Section 139a (4) sentence 1 SGB V[212]), partly implicit legal provisions regarding evidence-based medicine are gaining in importance.[213] To the extent that the law does take evidence-based medicine into account, it confers upon it a significance that extends beyond the specialist context in which it arises. This specifically rationalised and formalised (procedural) maxim, which is tailored to dealing with knowledge that is constantly evolving, and which originally referred to the physician's own clinical

211 Cf. in general on the concretisation of the complex requirements of SGB V Francke/Hart 2008; see also Axer 2011, 203 ff.

212 See also Section 73b (2) no. 2, Section 137f (1) sentence 2 no. 3, (2) sentence 2 no. 1, Section 139a (3) no. 3 SGB V, where the focus remains on evidence-based medicine. See also Wigge 2000.

213 Cf. to Section 35b SGB V, for example, Deutscher Bundestag 2005, 8.

decision-making process, thus acquires a general normative significance.[214] The assessment of methods that rely upon big data must also orient itself towards these standards.

As the authority responsible for overseeing statutory health insurance providers (as federal-level institutions), the Bundesversicherungsamt (German Federal Insurance Authority) does not regard the personal data transferred by policyholders to their insurers via smartphones, fitness trackers, or similar electronic devices and applications as valid evidence of their participation in quality-controlled methods or procedures in the sense of Section 65a (1) no. 3 of the SGB V. For this reason, the storage, processing and use of such data currently fails to meet the standard of necessity in data protection law.[215] It also remains unclear whether, regardless of the questions pertaining to data protection law, more fundamental considerations speak against a system of health insurance premium adjustment based on m-health applications and the data collected in connection with them. On the one hand, there is the obvious danger that the principle of solidarity could be eroded: among the core principles underlying a system of statutory health insurance is that insurers are "blind" to individual morbidity risks. On the other hand, it is clear that the aspect of personal responsibility and prevention in the SGB V is gaining in importance.[216]

3.3 Regulatory options

The opening up – in the sense captured by the term big data – of the use of data collected in the course of patient care for other applications and for medical research raises new questions and problems, increases complexity and has the potential

214 Cf. Stallberg 2010, 6 ff.; also Hart 2000.
215 Cf. Deutscher Bundestag 2016, 9.
216 Cf. the contributions in Weilert 2015.

to lead to confusion. Traditional data protection law is, as we have seen, insufficiently adapted to these challenges. A strict application of the existing regulations would ultimately mean declaring big-data applications broadly inadmissible, both in Germany and at the level of the European Union. Meanwhile, speaking against this outcome is the fact that use of data collected in the course of patient care for medical research represents a generally welcome development, promising both to expand the latter's knowledge base and increase its effectivity to significant degree. This fact can and should be accounted for within the leeway granted by constitutional law. Corresponding to the necessary stabilising function of the law is the need for sufficiently flexible regulations, open to innovation, particularly in those areas of modern life characterised by dynamic and rapid development. Therefore, the regulatory framework must meet fundamental ethical and constitutional standards; at the same time, it must also be designed in such a way as not to hamper innovation.[217] This speaks against the establishment of a comprehensive precautionary principle. It furthermore can be read as an argument for the utilisation of more complex regulatory inputs and more adaptable regulatory strategies involving private law and the cooperation of the private and public spheres. Institutional safeguards, if necessary by means of official legal provisions, are required precisely where rules in the private sphere do not imply true self-regulation. Where possible, these should not pertain to the concrete content in question as much as to organisational and procedural arrangements. A corresponding form of guided private-sector regulation can represent an important component of a more comprehensively conceived, coordinated, cooperative and phased legislative process.

The aim here is not to submit concrete proposals to lawmakers regarding precisely defined changes to be made to the existing system of legal standards. Because of the multitude of

217 Cf. in this regard Roßnagel/Nebel 2015, 459.

big data's potential applications, it is already nearly impossible to arrive at a uniform, conclusive judgement regarding its legal admissibility in the healthcare sector. Certainly, no comprehensive and detailed regulatory concepts can be deployed which address all the problems described. Granted, in view of the recent, thorough changes to data protection law in the form of the GDPR and the new version of the BDSG, it cannot be conclusively assessed whether and how the new norms and mechanisms will function. Nonetheless, it should be clear from the discussion above that, at the level of their content, certain principles of applicable data protection law can hardly be reconciled with the concept of big data. In view of the enduring, constitutionally based importance of the principle of data protection, this cannot be allowed to lead simply to an accommodation in the sense of reducing the applicable scope of data protection law. The goal cannot be mere deregulation, but can only be a transition towards smarter regulation better adapted to the domain in question. A corresponding evolution of data protection law should neither see the dissemination of data as inherently problematic nor should it downplay the real dangers associated with it. What should be sought instead is a concept for regulating and actively shaping developments in this domain that can differentiate and take into account both the opportunities and the risks associated with big data: "In the information society, data protection takes on a new quality, but not simply because so much data is accumulated, used and stored, but above all because a new data protection concept has become necessary. Flows of information can no longer be seen only as a danger to individual rights; rather, information and its distribution must be seen as an opportunity, even as the basis for the realisation of human potential, and thus must be evaluated in terms of the possibility of participating in communication processes."[218] Accordingly, the discussion that follows serves above all to highlight the various options of

218 Hoffmann-Riem 2000, 55.

regulatory action available in a more formally oriented sense. However, within the overarching notion of a mix of regulatory instruments elaborated below, certain desiderata will be singled out as instructive examples.

3.3.1 Further development of existing laws

Of course, laws remain the central regulatory instruments of the democratic, constitutional state. In particular, any decisions that pertain to the individual's basic rights must still take place at the level of legislation. Here again, however, it is important to highlight the difficulties that continue to arise in connection with big data as regards identifying especially sensitive data in certain areas. At least in the long term, it might therefore be useful to dispense with regulations specific to certain areas – e.g. data protection provisions specific to the social insurance system – and instead set guidelines that are as uniform as possible.[219] However, already at this point, the plurality of lawmakers in the federal system must be taken into account. Furthermore – notwithstanding existing differences, which are not merely restricted to the level of terminology – any future regulatory strategy would also need to incorporate EU legislative authorities into the lawmaking process. In terms of the recalibration of the existing regulatory approach mentioned above, it would be especially worth considering whether the lack of concreteness characteristic of big-data applications (in the healthcare sector) can be compensated for by additional technical-organisational, material and procedural safeguards.[220] The data portability provision of the GDPR (Article 20) can serve here as a model. It not only sets the requirement to introduce certain technically compatible data processing systems, but is also intended to encourage the use of

219 Similarly, Fetzer 2015, 778.
220 Cf. Weichert 2014b, 835.

interoperable formats, thus contributing to an emerging competition in the development of new technologies that are more conducive to data protection.[221] For example, it is important to ask in this sense to which degree the challenges arising from the continuous decontextualisation and recontextualisation of data could be managed by complex consent models, the clearer assignment of responsibility and (technical, but normatively required) options for data providers to exert ex post influence on the use of their data. As an example, the following approaches should be considered:

Especially with regard to the requirement for consent, it has become clear that the deadlocked and dysfunctional provisions of data protection law are in need of systematic improvement. Instead of a single, uniform solution, a more refined, discriminating concept is needed that can give space to the peculiarities of this regulatory domain and the preferences of those affected.[222] The possibility of introducing modified consent requirements, which has already been the topic of extensive discussion in the field of bioethics (see chapter 4)[223] and has at least begun to be implemented in the healthcare sector (see section 3.2.2)[224], could be part of the solution, as could technical concepts referred to under the heading of the "consent assistants."[225]

In addition, however, it must also be remembered that it is fundamentally permissible under data protection law to employ not only opt-in models but also opt-out models, i.e. in

221 Cf. Paal/Hennemann 2017, 1701.

222 Blanket consent (only) has a certain significance in the context of criminal law: With regard to Section 203 StGB, it has an exclusionary effect, even if it does not meet the data protection requirements of Section 4a (1), (3) BDSG. See Heimhalt/Rehmann 2014, 204 f.

223 See also, for example, Budin-Ljøsne et al. 2017; Kaye et al. 2014; Steinsbekk/Kåre Myskja/Solberg 2013; Wee 2013; Williams et al. 2015.

224 See also Richter/Buyx 2016, with further references. The Opinions of the National Ethics Council (Nationaler Ethikrat 2004) and the German Ethics Council (Deutscher Ethikrat 2010) on biobanks also indicate that a narrowly defined consent concept is problematic. See, similarly, European Commission 2012c, 57 f.

225 See, most recently, Riechert 2016 for details, in particular 24 ff.

order to collect and use data on the basis of legal standards of permission. For in terms of data quality and the importance of being able to draw upon datasets that are as complete as possible (representativity), an absolutised consent model can lead to problems. Firstly, because big-data analyses rely upon access to data that was originally collected for other reasons, a strict and comprehensive consent requirement would massively limit the amount of available data. Secondly, because only parts of the affected data are collected, a comprehensive consent requirement under data protection law would generate certain distortions. In any case, this could reduce the positive effects hoped for from big-data applications. There thus appears to be a contradiction between the ideal of voluntary data transfer – at least according to the basic principle – and the simultaneous pursuit of as full and representative a body of data as possible.[226]

In particular, Sections 22 and 27 of the BDSG, new version, contain corresponding data-use authorisation provisions, which, in compliance with European law, diverge from the prohibition on processing special categories of personal data laid out in Article 9 (1) of the GDPR. According to Article 9 (2) of the GDPR, EU member states may legislate exceptions to this prohibition. Accordingly, Section 22 (1) of the BDSG, new version, establishes the exceptional conditions under which the processing of special categories of personal data is permissible, without directly excluding any measures specific to this domain that may result from Article 9 (2) of the GDPR or be enacted on the basis of the GDPR as a whole.[227] This provision thus presupposes a complex weighing of interests (see chapter 1) which it supplements with procedural and organisational requirements (see section 2.1). Similarly, Section 27 of the BDSG also permits in particular the (further) processing

226 Cf. Augsberg 2016, which is exemplary with regard to the Transplantationsregistergesetz (Transplantation Registry Act).
227 Cf. Deutscher Bundestag 2017, 93.

of data for scientific purposes, on the basis of a comprehensive weighing of interests. In this way, the provision contained in Article 5 (1) (b) of the GDPR, according to which further processing in the service of scientific or historical research or for statistical purposes is not considered incompatible with the original purpose of data collection and can thus proceed on the legal basis applicable to the original processing of this data, is extended to special categories of personal data.[228] These regulations concerning the authorisation of data use must then be taken into account in the context of the protection of personal data under civil and criminal law, as well as within the framework of the special protections afforded to patient data.

The securing of data quality, as described above, can be supported by further legislative measures. This already applies at the level of data acquisition. The goal here is to counteract certain biases that, as we have seen, can arise as an undesired byproduct of consent requirements. Ensuring that datasets are meaningful[229] and free from discrimination[230] is one of the general challenges presented by big data. On the other hand, the degree to which manufacturers and service providers themselves critically appraise and verify the reliability of their data analysis is more than questionable. At least in part, these actors are also hardly capable of conducting a robust verification of the reliability and validity of their data. This is the case, for example, when consumers or patients themselves input data when using an app or influence through their behaviour the process of measurement and analysis. Clearly, this is especially problematic when a therapeutic strategy is linked to this data's analysis, but it also presents specific problems when such data is used for research purposes. Going forward, it is thus important to search for ways in which government can better support, but also, where necessary, oversee and

228 Cf. ibid., 98.
229 See, for example, Kim/Huang/Emery 2016.
230 Cf., for example, Barocas/Selbst 2016, 673.

safeguard, quality assurance mechanisms within science and industry. Should such measures prove to be inexpedient, more rigid legal requirements could be established in the medium to long term.

3.3.2 The regulatory function of private law

Alongside the readjustment of data protection laws, private law will also have an important role to play in the future development of data protection. The regulatory potential[231] that exists here can be illustrated with reference to three areas: consumer law, liability law and the rules governing the attribution of data and the power to determine its use (i.e. the "ownership" of data).

Consumer protection law is of twofold interest here. Firstly, like data protection law, it is based on the realisation that existing information asymmetries can lead to misuse. Situations in which the information available to the individual is limited can restrict the individual's constitutionally protected private autonomy and weaken the position from which they are able to act. These kinds of disadvantages are meant to be mitigated by consumer protection law, which concerns contracts. In their participation in economic life, consumers shall be protected from fraud and unfair advantages. Furthermore, data protection law serves to ensure that the consumer, as a legal person, is generally empowered to make optimal market decisions.[232] Secondly of interest here, in comparison to data protection law, are the significantly more complex combinations of information requirements, revocation options and mandatory contracts in the realm of consumer protection regulation. Thirdly, there are overlaps, to the degree that it is

231 See Binder 2012, 50 ff.; Hellgardt 2016; Poelzig 2012. See also Buchner 2006, especially for data protection.
232 Cf. Tamm 2011, 19.

discussed in contract law, regarding whether and to what extent the provision of personal data can be recognised as a form of monetary compensation.[233]

In Germany, liability law (which is understood here in terms of liability for the misuse of data) is characterised on the one hand by the negative right to injunction against misuse, independent of the question of fault, and on the other hand by the entitlement to damages, which presupposes fault in accordance with general civil law, provided no special liability statues are applicable (e.g. product liability statutes or medical device statutes). Thus, for example, in the case of medical devices[234], liability in the test phase must be distinguished from liability during the use phase; furthermore, there is a diverse range of potentially liable concerned parties to identify (hospitals and/or doctors, sponsors, manufacturers, suppliers, distributers, and notified bodies).[235] Here, it would be prudent to look for new ways to more effectively enforce existing liability claims or to establish new and more effective sanctions for the unauthorised collection and use of data. The GDPR already represents a significant step forward in this respect. In addition to criminal penalties (Article 83, 84 GDPR and Section 42 BDSG, new version) and fines (Article 83 GDPR, Section 43 BDSG, new version), it also entitles any affected person to claim damages from controllers and processors in case of unauthorised data use; this entitlement does not presuppose fault and encompasses both material and immaterial damages (Article 82 GDPR). A controller or processor is exempt from this liability only "if it proves that it is not in any way responsible for the event giving rise to the damage" (Article 82 (3) GDPR). In light of the by now undeniable fact that data has the potential to create enormous economic value, the efficacy of such liability-based regulatory approaches is beyond doubt.

233 Cf. Fezer 2017a, 100; Specht 2017, 763 ff.
234 Cf. for example, Reich 2014.
235 For details, Ortner/Daubenbüchel 2016, 2921 ff.

However, neither the GDPR nor German liability law are able to reliably compensate for the fact that, while big-data applications are admitted on the basis of their potential to serve the public interest, the economic benefits – at least initially – go to the data user, while the risks are borne by the data provider. When, for example, someone is reidentified through a company's use of their data, which they have authorised, but as a consequence is denied or loses a desirable insurance coverage, this does not justify a claim for damages under Article 82 of the GDPR, because the data use was authorised.[236] Material or immaterial damages suffered by this person as a result of the authorised use of data are not covered by Article 82 of the GDPR nor by German liability law.

The question of how and to whom the economic potential of data can be allocated is therefore of particular interest. Especially in the areas of clinical diagnostics and medical research, major uncertainties surround this question. The commonly deployed concept of the "ownership" of data is the starting point for numerous conflicts. There are a number of people who (mis)understand themselves as the owners or co-owners of data, because of the fact that they made important contributions to the collection, analysis and/or linking of this data, and who consequently claim exclusive rights to its use. In contrast to this view, it is important to note that in the strict legal sense, data cannot be owned. Ownership can only apply to (moveable or immoveable) objects. Data has no corporeality and is therefore not an object. Applying the rules for property ownership to data is therefore not possible under applicable law.[237] Because it has not been sufficiently processed, data in itself also does not fall under the provisions of intellectual property law.[238] The

236 Cf. Kreße 2017, para. 18 ff.

237 Cf. for instance Ensthaler 2016, 3475 f.; Paal/Hennemann 2017, 1698; Wiebe/Schur 2017, 463 f.

238 Cf. Ensthaler 2016, 3473 f. Accordingly, it is only possible to provide protection for already existing data (but not for data yet to be generated) under database law, as contained in Sections 87a ff. of the Urheberrechtsgesetz (Act on Copyright and Related Rights). See Wiebe 2017 for details.

way that the right of informational self-determination is legally designed analogously to property rights changes little in this regard, as does the new right to data portability contained in Article 20 of the GDPR. While the latter does de facto impact the allocation of rights to data, it does not touch on the question of ownership.[239]

Creative policy and legal approaches will be needed if one considers the current legal situation to be unsatisfactory and wishes to develop it further. Interesting strategies are already being explored with regard to the especially pressing matter of determining to whom a person's data that has been processed by researchers and/or companies "belongs", with all of its economic and scientific relevance. It is being analysed, for example, whether and to what extent data protection law already gives the persons affected, i.e. the persons whose data is processed, a legal position which at least approximates ownership of personal data – in the sense of a transferable exclusive right. On this basis, various ways to design legal ownership of personal data and to incorporate this design into the existing framework of property ownership are being examined.[240] Elsewhere, the creation of an "original intangible property right sui generis that is applicable to the data that citizens generate through their actions"[241] is considered, which should give "the user, as the data producer, defence and property rights proper to ownership."[242]

3.3.3 The possibilities of cross-border regulation

Particularly when it comes to big data, the required normative framework has a strong international component. The traditional modalities of control described above are all confronted

239 Cf. Keßler 2017, 591 with further references.
240 See in this regard Thouvenin 2017.
241 Fezer 2017a, 99 ff.
242 Fezer 2017b, 3 ff.

with the problem of responding with territorially bound legislative instruments to a phenomenon of data sharing and allocation that is, by its nature, tied up in digital networks and not restricted to national borders.[243]

Internationally speaking, however, prevailing data protection laws differ widely, which poses special challenges for affected parties as well as regulatory authorities. Data flows are for the most part cloud-based. The servers and firms that collect, forward and evaluate data are also to a large extent not based or physically located in Germany. By contrast, according to the so-called territorial principle, (only) such responsible parties fall within the scope of German data protection law that collect, process or use personal data in Germany. Corresponding provisions are also contained in the special data protection rules for media and/or telecommunications companies (Section 11 ff. Telemediengesetz [Telemedia Act]; Section 91 ff. Telekommunikationsgesetz [Telecommunications Act]). Here, the GDPR brings a significant improvement (especially Chapter V: Transfers of personal data to third countries or international organisations, Article 44 ff.).

Among other things, the cross-border dimension also places special demands on cooperation between regional data protection authorities and requires complex international coordination.[244] A corresponding instrument at the EU level is the European Data Protection Supervisor (EDPS). The institutions cooperating with the EDPS include Europol (law enforcement), the Schengen Information System (SIS, a database used for police cooperation and border management), EURODAC (a fingerprint database of asylum seekers and irregular immigrants into the EU) and the Council of Europe, to name but a few. Questions on health data are the subject of cooperation between the EDPS and the OECD (Organization for Economic Co-operation and Development). The latter, currently

243 International Bioethics Committee 2017.
244 Cf. for example Lewinski/Herrmann 2016 for details.

an association of 35 member states, has adopted guidelines on the handling of personal data in the private sector, which also include recommendations on health data.[245] The EDPS also participates in networks to support regional initiatives whose aim is to strengthen data protection worldwide. These include the Global Privacy Enforcement Network (GPEN), the Asia Pacific Privacy Authorities (APPA), the French-language association of authorities responsible for the protection of personal data (Association francophone des autorités de protection des données personnelles, AFAPDP), and the Ibero-American Data Protection Network (Red Iberoamericana de Protección de Datos, RIPD).

As an immediate successor to the Safe Harbour agreement[246] between the European Union and the United States, the EU-US Privacy Shield[247] came into force on 1 August 2016. These agreements are similar in terms of the regulatory technique they employ: with regard to the privacy obligations of American companies, a self-regulatory system is established whereby these companies voluntarily commit to comply with certain, specified privacy principles. Compared to the Safe Harbour agreement, the Privacy Shield has strengthened the rights of affected parties, insofar as participating companies and US authorities are now obliged to set up complaint bodies, and affected parties can now direct their complaints via their own national data protection authorities to the United States Department of Commerce. The agreement also stipulates

245 See OECD 2013.

246 See Commission Decision of 26 July 2000 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequacy of the protection provided by the safe harbour privacy principles and related frequently asked questions issued by the US Department of Commerce (OJ 2000 L 215/7).

247 See Commission Implementing Decision (EU) 2016/1250 of 12 July 2016 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequacy of the protection provided by the EU-U.S. Privacy Shield (OJ 2016 L 207/1). Cf. Towfigh/Ulrich 2017, para. 44 ff. with further references; Determann/Weigl 2016, 811 ff. with further references.

regular compliance and effectiveness reviews by the European Commission.

Regardless of the efforts to harmonise data protection legislation and improve cooperation and coordination in the field of data protection, from the perspective of those affected, the primary concern is the cross-border enforcement of their rights and their recognition by a foreign legal system, i.e. questions of international procedural law and international private law. The rules governing the enforcement of personal data protection rights are largely harmonised across the European judicial area; worldwide, however, the picture remains far more varied. Even where norms have been harmonised, there are many practical obstacles to effective prosecution.

3.3.4 The supplementary function of non-government control instruments

Regardless of the sometimes sharp criticism[248] aimed at such approaches, the example of Privacy Shield cited above illustrates that it may be appropriate to incorporate non-government and co-regulatory control mechanisms[249] into data regulation, particularly in view of the territorial limitations of classical legal controls, but also in light of the specific dynamism and volatility of this particular regulatory domain. Accordingly, an industry association in the digital sector proposes "ensuring transparent processes, confidence, and safety […] by means of certification with a recognised data protection seal (e.g. ePrivacyseal, the ULD seal)."[250] In addition to more general recommendations[251] regarding best practice, concrete proposals

248 See in general Lewinski 2016, 414 ff., 418; especially with regard to data exchange in medical research, see Molnár-Gábor/Kaffenberger 2017.
249 Cf. Zook et al. 2017 for a corresponding code of conduct for scientific research.
250 Bundesverband Digitale Wirtschaft 2017, 15.
251 Cf. Zook et al. 2017.

for more complex regulatory arrangements have also been made, for example by the Rat für Informationsinfrastrukturen (German Council for Scientific Information Infrastructures, RfII).[252]

Expanding the perspective in this manner to include industry self-regulation and/or co-regulation has a certain tradition in the context of data protection law.[253] The effectiveness of such an approach must not be overstated,[254] nor may it be employed as a "fig leaf" or substitute for necessary provisions in areas concerning sensitive questions of basic rights. However, in combination with such provisions, it can contribute to improving the quality and trustworthiness of applications and service providers, while aiding data providers in their decisions. Examples include explanations that go beyond obligatory information in describing how an app or a research program works, what it is capable of and what it hopes to accomplish, how data is worked with, how it is evaluated, whether and how data is shared, the extent to which users can themselves extract this data from the system and share it with others, or make modular decisions about the further use of their data, etc.

It should be noted at this point that government and private-sector regulation are not necessarily to be thought of in isolation, but rather can be interlaced. Thus, Article 42 of the GDPR allows for the certification of concrete steps in the processing of data. Furthermore, on the basis of Article 27 of the 1995 Data Protection Directive[255], the European Commission has already pressed for industry self-regulation to be promoted in the form of codes of conduct. Therefore, it is hardly surprising that the decisive impulses for a corresponding regulatory realignment emanate from the EU level,

252 See Rat für Informationsinfrastrukturen 2017.

253 See in particular the contributions in Wright/De Hert 2016; see also Martini 2016, 354.

254 Cf. for example Gellman/Dixon 2016.

255 Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (OJ 1995 L 281/31).

or that, especially for m-health applications, there exists today a relatively complex self-regulatory framework that combines compulsory provisions of the GDPR with a code of conduct, which in turn is backed up by special institutions and sanctioning mechanisms.

The initiative behind this co-regulatory measure can be traced back to the public consultation on the Green Paper on mobile health[256] and the doubts that emerged at the time regarding the ability to ensure an adequate standard of data protection.[257] A "Code of Conduct on privacy for mobile health applications"[258] was drafted on the basis of a comprehensive standards-setting process, moderated by the European Commission and incorporating a variety of stakeholders, and was formally submitted for comments to the Article 29 Data Protection Working Party on 7 June 2016, in accordance with the option to draw up such codes of conduct as foreseen in Article 27 (3) of the Data Protection Directive. In terms of its scope, the Code of Conduct covers m-health applications in which individual health data is processed and disclosed to third parties for medical purposes. While, on the one hand, apps are not regulated as long as they make use of mere lifestyle data that has no "clear and close" connection to the health condition of the person affected, on the other hand, the Code also does not consider applications that are already certified as medical devices. In terms of content, the Code, in keeping with its original history, aims above all to render fundamental data protection requirements and principles – informed consent, the obligation to adhere specified purposes, data minimisation – operable across new fields of application, especially from the perspective of app developers. Special attention is devoted to privacy by design and privacy by default, which are technical aspects that precede the actual use of data and must already be

256 See European Commission 2014.
257 See also Gassner 2016, 114.
258 See European Commission 2016.

considered in the development phase.[259] It must also be mentioned that the Code is accompanied by extensive institutional arrangements, in particular comprehensive regulations for the enforcement of the provisions it contains.

3.4 Conclusion: Statics and dynamics of the legal framework

Balancing the risks and opportunities posed by new technologies has rightly been described as probably the greatest socio-political challenge of our time.[260] However, the difficulty of the task must not lead one to conclude that it is insoluble. Instead, the task at hand is to determine, by examining the regulatory deficits that currently exist in this field, those fundamental ethical requirements that are indispensable and that must be upheld even under the conditions of big data. The goal must be, on the one hand, to be able to make use of the great possibilities offered by big data not only, but especially in the health sector. On the other hand, however, it is not simply a matter of maintaining the current level of data protection. On the contrary, it is a question of developing a core, guiding principle that is equal to the particular challenges posed by digitisation, towards which a policymaking and regulatory concept can then be geared.[261] However, before these thoughts on data sovereignty can be further elaborated and developed (see chapter 5), a detailed investigation of the ethical foundations of this issue (see chapter 4) is required.

259 Cf. also Becker 2017, 175 ff., who pleas for a "right to products free of data collection."
260 See Polonetsky/Tene 2013, 26; similarly, Roßnagel/Nebel 2015, 459.
261 See similarly Krüger 2016, 190: "Legal framework conditions are needed that ensure an appropriate balance between the data sovereignty of the individual and legitimate economic interests in the use of personal data."

4 ON THE ETHICS OF BIG DATA AND HEALTH

In light of the developments that big-data applications have undergone in the health and healthcare domain, major changes in the ways we experience health and illness at the level of individuals, science and society are in some cases already underway, while in others they can be predicted without much difficulty. Additionally, as chapter 3 has sought to make clear, existing legal regulatory regimes exhibit obvious shortcomings. In the following, it will be described how fundamental ethical norms and concepts are being challenged and modified by the technical applications being employed in biomedical research and practice and in other health-related fields.

On the one hand, the dynamics that we can expect to unfold in these areas will impact patterns of ethical orientation that evaluate and establish norms in relation to the role, function and position of the individual user of big-data applications. The relevant terms in this respect include freedom and self-determination, but also privacy and intimacy, sovereignty and power, as well as beneficence and non-maleficence, which will all play a role in designing normative models for the protection of the individual in the context of intensive data collection and exploitation. All these terms express the desire and entitlement of individuals not simply to become the object of data streams and the algorithms applied to them, but rather to retain a sufficient degree of control, sovereignty and power over their own data, or at least to be able to rely on third parties to protect their interests.

On the other hand, big-data applications in the healthcare sector impact upon measures of social orientation such as justice and solidarity. These concepts illuminate the fact that people have rights to certain things and that they care for each other in reciprocal relationships that are constitutive of communities. In light of the ways that the actually or supposedly

increasing exactitude of digital surveillance in all areas of life[262] affects the self-image of the individual, such bonds could weaken or even dissolve; either way, they will change significantly. For example, the solidarity principle underlying the system of statutory health insurance in Germany could be called into question, not to mention the possible consequences that could result in the domain of private health insurance. Conversely, it is conceivable that new, normative patterns of solidarity could also arise, for example, in the willingness to provide personal samples or data in order to promote medical progress from which one does not directly benefit.

Against this background, it will become more difficult to ascribe moral responsibility to individuals or institutions under the conditions of big data. Above all, it must be asked who can be held accountable not only legally but also morally if algorithms and the consequences of their sophisticated abilities of pattern recognition are not comprehensible, if machines "learn" to pass judgment on people on the basis of vast quantities of data and if people make decisions on the basis of the evaluation patterns that result from this analysis.

The following will not only outline a basic understanding of these concepts and their possible transformation in the face of the increasing use of big-data applications in medicine, research and other health-related areas. In this context, the vulnerability of affected individuals and groups that is associated with these changes will also be examined. Against the background of these analyses, chapters 5 and 6 will develop the outlines of an appropriate design and regulatory concept that takes up the fundamental concerns of the aforementioned normative orientation patterns, and attempts to implement them theoretically and practically in the current digital age.

262 Cf. Kucklick 2016.

4.1 Freedom: authorship and self-determination

In its landmark census ruling, the Federal Constitutional Court highlighted informational self-determination as an essential aspect of the general right of personality.[263] As chapter 3 makes clear, the set of developments in information technology that are summarised by the term big data, and which were discussed in chapter 2, have rendered existing data protection law, which was designed to preserve this right, in need of reform in several important respects. Data protection law can no longer sufficiently guarantee informational self-determination. Against this background, the task at hand is to redefine, or more precisely define, the normative core content of this protected good.

4.1.1 Authorship of one's actions

The term freedom is used to mean many different things and is deployed in myriad contexts, often with little precision. The distinctions developed below aim primarily to provide a conceptual reconstruction of human action, which is largely based on the experience of living and acting in the world. It is a fundamental human experience of action, presumably shared by all, that actors attribute to themselves and others authorship of certain events and states. A distinction must be made between this fundamental assumption and the concrete conditions associated with the exercise of authorship over action, which influence and constrain it to varying degrees. These concrete, personal conditions of action are summarised here under the term self-determination. The distinction between authorship of action (as a fundamental condition of freedom) and self-determination (as being dependent on more or less clearly perceptible circumstances) corresponds to a

263 Cf. BVerfGE 65, 1.

widespread characterisation of self-determination as the practical actualisation of freedom in an individual person.[264] This account of the concept of freedom covers a wide spectrum of normal action, ranging from decisions carried out in a state of full consideration to numerous everyday behaviours (such as smiling unconsciously when greeting someone) to unintentionally expressive acts (such as cheering by fans at a football match). All these actions are easily attributable to authors and are therefore fundamentally free in the sense intended here.[265]

The capacity for authorship of action is the basis of autonomy, i.e. the capacity by which people as agents are able to orient their actions according to maxims which they themselves set.[266] If human agents did not have this ability and were forced to

264 A detailed investigation can be found at Gerhardt 1999, 107-147.

265 The concept of authorship of action is, however, the subject of numerous controversies in more recent philosophical discussion. Controversial, for example, is the question of whether it alone designates an event that can be objectively determined in each case (the setting in scene of an action by an actor) or whether it also presupposes – and if so in what mode and to what extent – a subjective experience of action by this actor. Both do not have to converge side by side. (In certain borderline cases, for example those of schizophrenia, actors perceive their own actions as controlled by third parties; in other cases they perceive their actions, which are triggered solely by third parties, for example by certain brain stimulations, as genuinely their own, self-regulated actions.) In addition, numerous individual questions are controversial, both (1.) with regard to the objective foundations of an attribution of action and (2.) with regard to the phenomenology of the subjective experience of action and finally (3.) with regard to the relationship between these two types of conditions of authorship of action that may be deemed necessary. See Roessler/Eilan 2003 and Hyman/Steward 2004. However, for our purposes, these questions do not require discussion. Rather, we concentrate on the "normal cases" of human action, which, despite all their diversity, regularly exhibit *both* constitutive elements of authorship, i.e. the objective element as well as the corresponding subjective element – to whatever extent, measure and proportion.

266 Maxims in the sense used here are the subjective principles of one's own actions with medium generality and reach; they encompass more than individual motives for action and less than one's entire conduct of life. (Examples: to help those in need on principle; to fend off legal violations regularly; to develop one's own abilities as far as possible, etc.) In the history of philosophy, the concept of the maxim (as a reference object of the categorical imperative) was primarily developed by Immanuel Kant. See Grundlegung zur Metaphysik der Sitten (1785), AA IV, 385-464 (Kant 1903, 420 f.). According to Kant, the actor sets maxims themselves by virtue of "pure practical reason" and therefore acts (and only to that extent) "autonomously" (themselves legislating). See ibid., 402, 431 and see also Gerhardt 1999, 406-413.

carry out actions solely in obedience of authority, tradition or external compulsion, or if actions did not permit any rational description other than one expressed in terms of causal, deterministic processes, be they natural (physical, genetic, neuronal) or psychological-social (specific to family or social position, economic), then to speak of freedom would be to ignore the core of its true meaning. For these naturally or socially given conditions do not rule out, but in a certain sense even presuppose, that the person establishes their own existence in relation to them (by, among other things, overcoming them or subjugating oneself to them). Of course, it is always true that people do this as part of their belonging to the external world – the world as it is described and explained, for example, by biology and physics – and thus do so on the basis of processes which are subject to and constrained by natural laws and regularities (including those taking place in the brain). However, this does not exclude a reasonably understood concept of personal freedom qua authorship of action, nor does it rule out sufficient self-determination in concrete cases of individual action.[267]

The characteristic of authorship of action either belongs to a being or does not (*contradictory opposition*).[268] If, with Immanuel Kant, one assumes that the actor sets themselves as the ultimate end of their actions and demands recognition for this, which they must also grant to other actors, and if one thus demands that an actor is never to be treated as the mere means for the ends of others (prohibition of instrumentalisation), then one can speak of a recognition of the dignity of such actors.[269]

---

267 The philosophical problem of freedom of will cannot be dealt with here even in a suggestive way (overview in Kane 2011). Over and above all the disputes between the numerous positions, there is a far-reaching consensus that a reasonable concept of human freedom is explicable for all sides (albeit with different requirements for this concept).

268 A static contradiction is one that, in the interest of avoiding a contradiction, allows only yes-no assertions (example: coloured – colourless), while a dynamic contradiction allows discrete or continuous transitions (example: black – white).

269 See e.g. Grundlegung zur Metaphysik der Sitten (1785), AA IV, 385-464 (Kant 1903, especially section 2, 427-429, 437) on the prohibition against instrumentalisation especially ibid., 438.

Irrespective of the question as to whether non-human agents (for example animals or "super-intelligent" machines) can also be considered the authors of their actions, the ethical considerations elaborated below are limited to humans.

4.1.2 Self-determination and consent

Freedom, in the sense of the concrete realisation of authorship of action, shall henceforth be referred to as *self-determination*, in accordance with the distinction proposed above. Self-determination can be exercised more or less, or even not at all, by the authors of action (*polar-contrary opposition*)[270]. In this sense, self-determination is realised in different forms.[271]

In one sense, self-determination refers to the *ability* of a person to shape their life, at least on the whole, in accordance with their idea of how they wish to live. The degree to which this ability actually exists can vary. Below a certain minimum requisite threshold, one can also no longer speak of self-determination in the sense of a mere disposition, for example in the case of small children or those suffering from serious mental illness. Exactly where this boundary is to be drawn, however, cannot be determined in abstract, general terms, but rather depending on the specific context of action and the specific goals pursued.

In a further sense, self-determination refers to *an actual state*: even those who possess the capacity for self-determination need not exercise it in their personal life, or at least not continuously, either because they are prevented from doing so by external circumstances (for example imprisonment or extreme need), or because they neglect to realise the ability they possess and largely surrender their conduct of life to the

270 See footnote 268.

271 The following remarks are based on the investigations carried out by Feinberg 1989, 27-51.

directives of others. In this case, one can speak of the absence of the actual state of exercised self-determination. A number of elements can be understood as the criteria of a self-determined conduct of life in this sense, which any life considered practically self-determined would have to exhibit, such as sufficient independence from the decisions of others, sufficient authenticity of one's own decisions, and sufficient self-control of the actions by which such decisions are carried out.

In a final sense, self-determination refers to an ideal, and idealised, *way of life*. Even those who, in the above sense, can unquestionably be regarded as (sufficiently) self-determined may nevertheless be more or less far removed from the ideal concept – perhaps their own – of the self-determined person, and from a corresponding way of life. If one accepts this ideal concept as in itself binding, one may derive from it the imperative to further develop one's own personality in pursuit of it. As an ideal, the goal does not have to be attainable. Rather, it should be understood as a regulative idea for the orientation of one's own, actual way of life.

These forms of personal self-determination must be distinguished from the *legal* protection afforded their exercise. Constitutional as well as non-constitutional legal norms guarantee in many respects the ability to act according to truly personal maxims, reasons, wishes and interests. However, the scope of their legal protection is not congruent with the objective scope of these forms of personal self-determination (as ability, condition, or ideal), but rather usually extends beyond it. Most people make at least some of their decisions according to other people's standards, public fashions or collective moods, and some make most of their decisions along such lines. This may not come close to satisfying the criteria of personal self-determination set out above. At the same time, however, it also counts – up to the point of being coerced or based on manifest deception – as legally self-determined behaviour and thus falls under the protection of the relevant standards. Actions that are externally motivated or determined in this way are thus

still wholly legally binding. Anyone who in this manner acts dependently but nevertheless, from a legal point of view, with self-determination, can still legally refuse to tolerate any (paternalistic) interference on the part of third parties.

Of course, this difference in scope between truly personal self-determination on the one hand and legal self-determination on the other has a downside, the significance of which is obvious precisely in the context of big-data technologies. Insofar as external influences on others' decisions do not exceed the limits of their legal self-determination (i.e. insofar as they exert their influence through more subtle means than deception or coercion), they are in principle legally admissible. Nonetheless, these influences may well substantially endanger the ability of those they target to determine their own actions themselves or to lead a life based on self-determination; certainly they may undermine the guiding criteria of self-determination as an ideal. This applies in particular to strategies of influence whose effects lie below the threshold of their targets' perception. But it is precisely this that specifically characterises forms of influence based on big data. As "nudges"[272] operating on a mass scale in economic, political, religious or ideological matters, they may not at all be noticed by the users of online services; however, precisely for this reason they may be powerfully effective in a way that crosses the threshold of collective manipulation but does not yet violate any legal prohibitions. This raises the question as to whether there is a need for new forms of legal protection against such new and subtle ways of undermining truly personal self-determination (see section 3.1).

272 The term "nudge," introduced by Richard Thaler and Cass Sunstein, stands as a collective term for measures intended to redesign a so-called decision architecture (Thaler/Sunstein 2008). The latter is modified in such a way that a certain desired behaviour becomes simpler, more attractive, more obvious, etc., without at the same time denying other options for action. (See section 4.3.)

One cannot waive or delegate the fundamental ability to make decisions and act; however, self-determination is a different matter. The form and degree of its exercise are of considerable practical importance. One can, for example, delegate one's right to self-determination to someone else as a precaution in certain contexts and for precisely defined purposes. This ability to delegate becomes particularly relevant for individuals whose self-determination is restricted in terms of their capacities or abilities, e.g. individuals who must take precautions in cases of dementia, which may reach a stage in which they cannot be said to exercise self-determination as a right even in the sense outlined above. If necessary and possible, abilities that are lacking must then be compensated for, e.g. by a representative appointed directly by the affected person (mandated self-determination), by a representative elected collectively by a group including the person affected (representative self-determination) or by a representative appointed for the person solely by others (advocatory or assisted self-determination). As a form of care, the "soft" paternalism that is rendered legitimate in this case is not necessarily the antithesis of self-determination, at least not if the wellbeing of the person affected, as they themselves understand it, is its determining criterion. For its part, paternalism towards children can also be justified if the person acting on behalf of the child always does so in the child's best interests and does not undermine the child's dignity through their decisions.

In the context of big data, a number of instruments have been developed in recent years to facilitate the practical realisation of the notions of self-determination outlined above, including by means of delegation. These include, for example, various ways for individuals to consent to the collection, storage, and further utilisation of their biomaterial and any associated personal data, as occurs, for example in large biobanks

where medical research is conducted.[273] In contrast to medical research carried out in a clinical context, when material and data is stored in biobanks it is usually not possible to predict which future research questions will be put to it. Furthermore, biobank-based research yields its best results when materials and data are shared with other researchers and institutions and linked with other datasets, none of which can be predicted fully in advance; this is also fundamentally true of most research based on big data. As a result, the classical model of informed consent used in clinical research, in which consent is closely bound to a clearly defined purpose, and in which research participants are informed in detail regarding research goals, questions, risks and the further use of results, is inexpedient in the context of biobanks.[274] To ensure practicability and effectiveness in this context, permission to use material and data must be more indeterminate and thus broader. At the same time, it must be ensured that this permission is based on a truly self-determined decision on the part of the research participant.

One way of addressing this challenge is to simply obtain blank consent, which has been extensively relied upon in the US and UK. This involves the participant consenting, at the time their biomaterial is collected and their data stored, to the (undetermined) future use or transfer of their material and

273 In biomedical research, two types of biobanks have so far been the most common: first, mostly smaller, disease-specific biobanks in which patient data and biomaterials are collected for research into a specific disease or disease group, for example to enable disease-related genome research, although it is already difficult to determine an exact target even in these biobanks. The second type are population-based biobanks, in which data and materials from large populations or population cohorts are collected in order to carry out health-related research without close, e.g. disease-related, earmarking. (Often, however, a specific, purpose-related consent is given for each individual study within the biobank.) Increasingly, the boundaries between the different types are blurring. For example, as described in chapter 2, various previously disease-specific biobanks have merged into larger research repositories, or large healthcare institutions such as university hospitals collect all of their patients' residual materials and make them linkable with clinical data (healthcare-embedded biobanking).

274 See Deutscher Ethikrat 2010.

data; all further decisions regarding this use are then in the hands of the biobank operator. Whether an open-ended, one-time consent agreement of this kind, and the resulting de facto relinquishment of all future decision-making and control capacities, without any additional safeguards, can meet the threshold of self-determined agency is the subject of considerable controversy.[275] In the meantime, various other consent models have been established, with the aim of striking a balance between an unrealistically narrow restriction of purpose, on the one hand, and a one-time, exceedingly broad release of data and control, on the other.[276]

In this respect, the dynamic consent model has become well-known.[277] It seeks to do justice to the requirements of well-defined, informed consent by allowing this consent to be repeatedly and dynamically granted in relation to every project component and sub-project being pursued at a biobank. Participants thus have a reciprocal relationship with the biobank – usually via an online platform or by telephone – and regularly receive an options menu of sorts, which they can then use to decide how much information they would like to receive regarding further projects and whether they would like to participate in any of these projects by releasing their material and data.

The original model of dynamic consent was designed only to secure and facilitate personal consent, bound to or earmarked for specific future projects, and to streamline the information process. However, it was revealed that participants sometimes preferred to delegate at least some of these future decisions, for example, in the form of representative self-determination outlined above. The dynamic consent model has therefore been further developed and expanded. At present, a procedure known as cascading or meta-consent may be

275 Relevant here is Caulfield 2007.
276 Richter/Buyx 2016 provides an overview of the various models and ethical arguments in this debate.
277 See Kaye et al. 2015.

considered the gold standard.[278] On the one hand, it maintains the dynamic nature of consent while adding further choices, such as delegation options. On the other hand, a key decision is placed upstream of the entire process. Participants decide at the beginning – i.e. when donating material and data – which form of consent they generally prefer. This decision is based on detailed information regarding the overall objectives of the biobank and its main characteristics, i.e. its financing, its data protection and governance concept, any cooperations and transfer agreements it has with other institutions as well as the various options research participants have at their disposal for granting their consent. Participants are then asked to choose whether they would like to be informed dynamically about each future project and whether they would like to be asked each time to consent to their participation. Alternatively, they can consent to various broad categories of research without giving permission each time for individual projects; they can delegate future decisions to an expert panel, e.g. an ethics committee associated with the biobank or an independent supervisory body; in some versions of the cascading model, participants can release their materials and data for any and all uses; of course, they can also refuse to participate. The cascading model and the dynamic communication structures on which it is based make it possible for participants to switch between these various forms of consent, as well as to withdraw their consent for any further use. In addition, combinations of the various consent options are possible, such as delegation combined with a generally broad or comprehensive consent.

Implementing this cascading model beyond the existing pilot projects on a broader scale for biomedical research – and beyond – will involve significant organisational efforts by the institutions involved. On the other hand, it can be assumed that this model may most closely correspond to the goal of aligning the use, by others, of one's personal data with one's

278 See Ploug/Holm 2016.

own best interests and values – provided it is ensured that buttressing protective mechanisms, such as appropriate data protection concepts, can be made to function over the long term.

4.1.3 External parameters for the realisation of freedom

If self-determination is to be understood as the concrete, individual, practical enactment of freedom, then the characteristics of freedom identified by John Stuart Mill – namely individuality, authenticity and originality – are, not unlike the criteria outlined above, also forms of self-determination.[279] In order to judge each of these factors, the social context of the actor must also be taken into account. One can only realise one's capacity for freedom in a self-determined way if the necessary external conditions are present to a sufficient degree. These conditions can have a positive effect on self-determination, insofar as an individual is in a position to develop and implement their own life plans. A democratic and plural society presupposes that every individual may enjoy, without the need for further justification, sufficient scope in which to pursue their own life plans and intentions, as long as these do not illegitimately interfere with the rights of others. Self-determination in this sense, as an expression of the right to be left alone (often referred to as negative freedom)[280], is connected to the right to protected privacy and intimacy (see section 4.2), but does not mean mutual indifference amongst people. Rather, it is a personal condition and the fertile ground for developing one's own ability to lead a self-determined life guided by the standards of an ideal (often referred to as *positive freedom*). Conversely, there is a danger that the space of negative freedom will be emptied or thinned out if it is not oriented towards the development of positive

279 See Mill 1859.
280 See Berlin 1969.

freedom – as self-determination – in such a way that people are also committed to the communities and societies in which they live and to the questions of the common good that prevail there.

Against this background, to be free and able to act in a self-determined manner means, at the least, having the realistic ability to preserve and shape one's own identity and to answer for one's own actions before oneself and others. For this to be possible, especially in a plural and functionally differentiated society, necessitates constitutional standards that are fair and reliable and which apply regardless of an individual's reputation. Self-determination and freedom – both negative and positive, both internal and external – can only be realised over the long term within such a framework of just and legally secure institutions. This in turn presupposes that citizens, for their part, understand such a framework as essential to their own self-determination and thus for the realisation of their ideas of a good life, on both the individual and social levels. In this sense, positive and negative freedom as well as external and internal freedom are mutually dependent.

Thus, any restriction on the individual's external freedom must be examined very carefully to determine whether it entails a threat to internal freedom. In social ethics and political philosophy, for example, it is emphasised that institutions – at least public institutions and agencies – are only conducive to freedom if they are designed in a way that is transparent and guarantees fundamental access for everyone.[281] Furthermore, they must be designed in a way that empowers those affected by institutional decisions to understand these decisions to the greatest degree possible and, according to the circumstances,

281 A prominent example: John Rawls, for whom the openness of access to offices and positions for everyone under conditions of fair and equal opportunity is a constitutive element of the second of his (two) fundamental principles of justice, i.e. the so-called principle of difference (with which alone, according to Rawls, social inequalities can be justified). See Rawls 1972, 75 f.

to implement them in a sovereign manner. Finally, procedures must be in place that enable participation and control, so that, in principle, institutional structures and individual decisions made within their frameworks may be subject to revision. Despite all the practical obstacles to meeting such requirements in a complex, modern society – especially one that is, in the field of information, globally networked – it cannot be denied that such criteria represent standards for determining how individuals can preserve and defend their inner freedom such that they are not left to the mercy of external constraints.

4.2 Privacy and intimacy

The right of determination over one's own life, the right to be able to lead it accordingly and to orient it towards one's own motives, reasons, desires and choices, none of which may be forced, extends in principle to all areas of private and public life. In the public sphere, someone's self-determined conduct of life is bound in an intricate network of coordination with, and justification vis-à-vis, others. The public sphere is potentially always a medium of unwanted observation and control; it therefore has a policing or limiting effect on each individual's self-determined conduct of life. On the other hand, privacy refers to a sphere of life in which such unwanted control mechanisms and demands of justification are largely held at bay,[282] and to which the person concerned decides themselves to whom they wish to grant access.

The classical legal definition of privacy as the "right to be let alone" dates to 1890,[283] but the distinction between a private and a public sphere is both much older and familiar across

282 Even within the private sphere/legally protected space of one's own personal autonomy, of course, one cannot do everything one wants; here, too, the protective rights of those with whom one shares one's privacy continue to apply.

283 Cf. Warren/Brandeis 1890.

many cultures,[284] even if the definitions of these spheres and the boundaries to be defended, the grounds on which these definitions are justified and the threats to privacy that have been identified have all varied historically, with their relationships to each other constantly undergoing rebalance. The ancient cultures of the Greeks and Romans already distinguished between the public (*polis* or *res publica*) and the private household (*oikos* or *domus*). The same applies to large parts of the globe culturally shaped by Judaism or Islam.

Modernity has also been host to numerous traditions that tie in with and further develop this fundamental differentiation. Two lines of development are particularly relevant here: as regards the demarcation of the *oikos* from the *polis*, the private sphere is increasingly justified in terms of the logic of ownership. At the same time, it seeks primarily to define an area that is protected from state intervention. Thus, the modern idea of the rule of law is also associated with the idea of preserving for the citizen a clearly defined and reliable space within which they can regulate their personal affairs without outside influence (private autonomy). Of course, an early criticism of this conventional view was that the private sphere is threatened not only by the state, by also by a high degree of informal pressure on the part of society. Therefore, the classical definition of privacy as the right to be let alone refers not only to the state but also to other social forces such as mass media or communities of conviction, which can pose a considerable threat to privacy as a place where people can live their own lives beyond the influence of others.

This understanding of the private sphere – as a place where one can determine the conduct of one's life without being subject to influence or surveillance – also plays an important role in connection with a significant differentiation between two meanings of *oikos*. For even within the *oikos* or, in modern terms, within a family, delimited spaces of privacy that must be

284 Cf. ibid.

mutually respected arise amongst otherwise closely connected members. Against this background, someone secretly reading their partner's diary entries is, in the eyes of many, just as taboo as a parent monitoring the Facebook page or chat history on a mobile phone belonging to a teenager who does not wish to allow them access. At stake are spaces of intimacy reserved exclusively for those directly involved, who alone may grant access to a selected circle, if at all, in a self-determined manner. Even this is true only to a certain extent, for there are areas of intimacy whose voluntary, even deliberately provocative disclosure is often regarded by others as exhibitionism. Of course, what is considered intimate (or exhibitionist) is to a considerable extent culturally variable. Nonetheless, these terms' frame of reference is always connected in a special way to shame and humiliation. However, it is precisely this aspect of shame, to which privacy and, above all, intimacy are inherent, which always harbours the potential for abuse: if someone's intimate sphere is violated, or, in the worst case, if someone is subject to sexual violence, the disclosure of this abuse is highly fraught with shame, which is why potential victims often shy away, for reasons of self-preservation, from such disclosure. An awareness of this reverse side of privacy and intimacy does not negate the fact that they remain fundamentally deserving of protection, both as values and spaces, but it does caution against uncritically exaggerating their primacy. Privacy and intimacy do not constitute a context of legitimation for moral or legal violations of others.

The libertarian argument that only as extensive a private sphere as possible can guarantee personal freedom has been countered from, for example, a communitarian perspective by the assertion that, entrenched behind the bulwark of the private sphere, there reside primarily egoistical claims vis-à-vis legitimate collective concerns, which threaten to atomise society into individualised parts. In the context of big data, the post-privacy movement has also taken up this critique.[285] It

285 See Heller 2011.

demands the disclosure of all citizens' activities, in the interest of a transparent and therefore fairer society. Its exponents nonetheless forget that the urge to safeguard the possibility of intimacy and confidentiality protected by the private sphere derives from respect for the interpersonal communication that unfolds in immediate social proximity and, not least, for the profound vulnerability to which everyone is exposed by the fragmented and searching nature of the process of shaping their own way of life.

This suggests a normative justification of privacy based on its (social) anthropological significance. Only in the private sphere can close social relationships (with romantic partners, family members, friends, neighbours) and the conditions for the development of personal identity develop and develop anew. Privacy opens up spaces of intimacy and confidentiality where people can cultivate relationships in which they can genuinely and without pretence be themselves – shielded towards the outside, but open on the inside. It can be assumed that such intimacy and confidentiality is an indispensable precondition for the willingness to engage in the always uncertain and searching process of shaping one's life and thus to affirm and practice the always fragmentary aspect of any self-determined way of living. Only within this privacy can people preserve their unselfconsciousness and experience each other authentically without fear of a ruthlessly voyeuristic public.

Every human being's relationship with themselves would be dramatically altered by the knowledge that in principle they could, anywhere and at any time, be under surveillance. A safeguarded privacy protects against this possibility. However, controversy remains regarding how the concept of the private should be understood in detail. One possibility is a conception developed by the philosopher Beate Rössler, which limits it to a few functions but equips it with a strong normative claim. Rössler understands privacy as a basic condition of a self-determined way of life, but not only formally as a protection against outside interference and attack. Rather, according

to her model, privacy is realised above all in controlling access to direct contact for the purpose of establishing a close social relationship as well as to information concerning the intimate details of one's own life ("access to one's own person").[286]

According to Rössler's conception, the possibility of exerting this control is realised on three fronts: firstly, it must be possible to avail oneself of room for manoeuvre and decision-making; secondly, one must retain control over what others are permitted to know about oneself; and thirdly, there must also exist places of refuge within which and from which self-determination can be developed. Three forms of privacy can thus be distinguished: decision-relevant, informational and spatial privacy. With reference to the private nature of the decision-making process, this concept makes it clear that certain conditions and capabilities are necessary in order to be able to defend against outside interventions and attacks. Today, this dimension of privacy is often linked with the idea of psychological resilience.[287] Resilience, of course, does not simply occur naturally. It must be acquired and instilled through processes of education and development; it must be promoted and nurtured; and it often must be defended against the homogenising tow of social expectations of conformity and control.

Other approaches interpret privacy more cautiously, particularly in view of the highly different contexts in which it is used. Sociological and legal studies show that it is difficult to find a common denominator for an even rudimentarily shared understanding of privacy.[288] At the same time, even more restrained conceptions of privacy are designed to include differentiated taxonomies of possible threats to privacy in the field of digital communication, threats entailed by the collection, analysis and linking of data and information as well

286 Cf. Rössler 2001, in particular 23-26.
287 See, for example, Berndt 2013.
288 Cf. Solove 2007; 2008; 2011 and Solove/Hartzog 2014.

as deliberate interventions in the life and environment of the data subject.[289] It is precisely the possibility, mentioned in section 2.4.1, of originally anonymised data being deanonymised that is associated with an obvious potential for the violation of the private sphere. The growing capacity to link originally discrete items or sets of data for the purpose of revealing and determining new correlations – possibly without the knowledge and consent of the data subject – also has the potential to restrict privacy in the sense of the ability to exert control over data flows.

Nevertheless, people living in digital societies often do not completely reject the reuse of their data as a violation of their privacy, even if they have only involuntarily agreed to it, for example because of terms and conditions that grant no alternative. Research into individual and social attitudinal patterns confirms that, depending on context, people attach different and sometimes contradictory importance to their claims to privacy in both the analogue and digital worlds.[290] Furthermore, attitudinal patterns, expectations and experience prove to be key factors in shaping people's online behaviour: users of social media widen the circles of "friends" with whom they share private information, revealing more than they previously would have without the use of these media. People believe that by managing their "friends" they are able control the further use of their data.

Even the psychologically understandable pattern of breaking the rule, in certain situations, of normally sharing intimate things only with friends – e.g. revealing oneself through very personal remarks to a perfect stranger during a train journey – is likely to be more common in the digital age than it once was. Thanks to new digital data-linking possibilities and developments such as facial recognition technologies, which are slowly becoming established, even a few statements made carelessly

289 Cf. Solove 2007.
290 Cf. Acquisti/Brandimarte/Loewenstein 2015.

under the supposed cloak of anonymity can be assigned with ever increasing accuracy to a person, even years later. It should be immediately obvious that, under such already precarious conditions, the capture of health-related data and its subsequent processing and correlation using big-data-driven methods can produce extraordinary vulnerabilities, if, for example, the analysis of certain statements made on social media produces supposed indications of psychological abnormalities.[291]

Even if many people may suspect, despite possibly underestimating the risks, that in a digital society it has become impossible to completely control their own data trails, this does not mean that they are indifferent to how their data is used and reused. Depending on their preferences and context, people decide where, on the one hand, they wish to focus more strongly on data protection and control and hold data processors to higher standards of trust, and where, on the other hand, they are more likely to act in a relatively relaxed or even negligent manner. At the same time, for many people, it is important to have the option to change ad hoc their preferences and decisions in this regard.[292]

To the extent that the individual data subject's demand of exerting control over the use of their own data is increasingly perceived as illusory, the expectation that data streams will be treated confidentially by the organisations or companies accessing them gains in importance, in terms of satisfying whatever demands made of data handling that continue to be deemed realistic. In much the same way that many motorists do not understand the technical inner workings of their vehicles, but instead rely upon the reputation of a given brand as a seal of quality, people often simply assume that the online services upon which they most depend can assure a sufficiently robust degree of confidentiality when collecting and passing on their data and thus will protect their privacy. Clearly, such

291 See section 4.7.2 for an example of Facebook's suicide detection algorithm.
292 Cf. Lauss et al 2011.

a leap of faith is essential, for it would be simply unrealistic to expect that the normal, individual internet user should possess both the time and the technical competence to work their own way through the complex terms and conditions they encounter online, while fully grasping the consequences of the consent they are expected to provide. The ways in which, under the conditions of big data, information can be used, linked and disseminated are so complex and inscrutable that most people would be massively overwhelmed by the task of understanding them in any given context and providing their consent accordingly. They are thus left with the unrealistic option of completely withdrawing from online activity[293] or the pragmatic option of placing their trust in the data policies of a given organisation, which then can be said to represent the data and privacy interests of the individual in a fiduciary capacity.

Under no circumstances should individuals be relieved of responsibility in matters that they can decide for themselves, about which they can sufficiently inform themselves or for which they can delegate responsibility, as a precaution, to competent parties – especially in the area of health-relevant lifestyle products (see section 4.7). Nonetheless, in order to understand privacy as the fundamental ability to control one's own data, a conception of privacy protection needs to enter the focus of ethical reflection that revolves less around individual behaviour than around organisational standards of confidentiality.[294] It is this notion that must take priority. It, too, provides for some elementary moral and political norms and rules of privacy – especially the criterion of not having to incur damages through threats to one's privacy (see section 4.4). However, to simply uphold a fixed normative core of privacy will hardly suffice to meet the challenges currently presented by the evolving digital society in which we live. Instead, the task is

293 Cf., among others, Conley et al. 2012; Acquisti/Brandimarte/Loewenstein 2015; Floridi 2014.
294 Cf. Nissenbaum 2009 and Nissenbaum 2011.

to identify the contextual conditions specific to given privacy safeguards and threats both online and offline. Then the question arises as to which values and identities are most relevant for a person and in which contexts they render the protection of confidentiality especially necessary.

Certainly much of the information that a person reveals about themselves in a particular context is originally bound to said context. Thus, when driving or traveling by rail and employing a navigation system or online ticket service, indicating a frequently used route often translates into a real convenience for the user. However, in a digital society, and especially under the conditions of big data, information of this kind that is initially contextually bound can be linked, in real time, with other data and contexts in ways that are both myriad and unpredictable. Furthermore, there do not yet exist any effective, reliable and lasting means of ensuring that this data is forgotten or deleted.

It is therefore at the least negligent to assume in good faith that one has nothing to hide. All this proverbial assertion of innocence means is that one can currently discern no context in which their data might be used and in which the revelation of a given piece of information would be to their detriment. While such credulity is already problematic, under the conditions of big data it becomes definitively absurd. No one can know whether or not there are or will be contexts in which the use of certain information could prove at least undesirable and unpleasant, if not harmful, for the data provider. What appears unproblematic at a given time and in a certain context (e.g. merrymaking whilst on holiday) can produce knowledge in the course of a retrospective examination of one's lifestyle and behaviour (e.g. as part of a job application), the consequences of which can entail a severely negative impact on one's plans in life or those of the people around them.

Furthermore, the negative repercussions of self-induced heteronomy (see section 2.5.5) and/or informational self-endangerment (see section 3.1) are revealed with particular

clarity in connection with the ever-increasing number and variety of opportunities to disclose intimate details digitally. The self-induced provision of personal information evolves into a lifestyle that is largely dependent on external factors for affirmation and coherence. However, the more the possibilities of data linkage proliferate beyond the ability of affected individuals to foresee or control, the more imperative it becomes to legally and social-ethically establish online the standards of trust that people know and value from their offline lives.[295]

The question of how to protect and preserve privacy under the conditions of big data, and of the different ways in which this could be accomplished (data protection measures, stricter expectations regarding the confidentiality of data streams, etc.) affects not only individuals, but also groups. This becomes clear in the medical context when, for example, an individual's desire to learn about their genetic disposition affects or places a burden upon others. In this context, the "right not to know" is often cited, according to which adults may decline to receive information regarding potential genetic issues in their family even if a fellow family member could medically benefit from this knowledge. Questions surrounding how this right could be legally enforceable and at what point it comes into conflict with the rights of other family members to be informed are currently the subject of considerable legal and ethical controversy.[296]

The question of one's entitlement to privacy is rendered yet more acute by big data, as the potential to draw conclusions, by analysing massive quantities of data, regarding combinations of characteristics shared by other people now extends well beyond the family circle and can affect larger groups. Those affected are, often completely unbeknownst to themselves, grouped together by algorithms and assigned a certain target characteristic or label. In serious cases, such characteristics can have stigmatising, discriminatory or exclusionary

295 Cf. Conley et al. 2012.
296 Cf. Duttge 2010; Koppernock 1997, 89 ff.; Dorniok 2015 and Schroeder 2015.

repercussions. Because of mere statistical correlations with health issues (see chapter 2), certain aspects of people's lives, specific behavioural patterns or demographic traits, which the person concerned can do little to change, can be given a socially negative classification or be labelled risk-prone by certain interested parties (such as health or life insurance companies; see section 4.6). The correlations that big-data analyses reveal, on the basis of which groups of people who share combinations of features defined as precisely as possible are supposed to be identified, still do not necessarily justify simply assigning all individual bearers of such features to these groups, since further characteristics that are not recorded in the analysis can have a considerable influence on the individual probability of exhibiting the respective target characteristic.

In contrast to the classical medical situation, in which the process of ascertaining such combinations of features and communicating the risks associated with them can be carried out in strict confidence, this is hardly possible under the conditions of big data, as publicly available and/or – at least at first glance – not apparently health-related data points are increasingly drawn into the analysis. Not only might this mean that the individual can neither prevent nor understand their assignment to a certain risk profile; it also means that, on the basis of the combination of features at hand, third parties can also easily undertake or retrace such an assignment. This exacerbates the potential threats to privacy, as they no longer potentially affect only individuals or small (family) groups, but increasingly also larger groups at the societal level, to which certain risk profiles are assigned on the basis of combinations of features extracted by big-data analysis.

4.3 Sovereignty and power

As a stronghold of personal freedom over the course of one's life, the private sphere is essential to the exercise of

self-determination. If privacy consists in controlling access to all the places and accomplishments that belong to one's life, this must apply both offline and online, especially as the relentless digitisation of our human world is causing these spheres to increasingly merge.[297] The collection and subsequent evaluation, processing and dissemination of personal data always constitutes an invasion of privacy that requires legitimation.[298] This understanding of privacy is associated with an understanding of self-determination in the special sense of actively shaping one's self, a state of empowerment to which the concept of sovereignty has come to refer in everyday language and that can be further developed into the concept of data sovereignty (see chapter 5).[299]

Of course, the concept of sovereignty is by no means unproblematic and thus requires further exposition. On the one hand, it contains an emphatic tone: as a term invested with a high moral value, it signals a claim to unconditional regard that can hardly be called into question. For many, personal sovereignty is the basis and goal of a successful, self-determined life. It is thus not surprising that the term has found the most diverse applications across numerous areas of life – from consumer or customer sovereignty[300] to patient sovereignty in care contexts[301], and from food sovereignty[302] to seed sovereignty[303] (the unimpeded access to seed variety). Similarly, in the field of digital media, sovereignty appears in a number of

297 Luciano Floridi coins the term "onlife" to make it clear that online and offline life can hardly be distinguished anymore. Cf. Floridi 2014.
298 Cf. Pillay 2014.
299 Cf. Friedrichsen/Bisa 2016.
300 To a certain extent, it forms the normative core of the at least methodological individualism contained in (neo-)classical political economy (cf. Samuelson 1975).
301 "Patient sovereignty refers to the possibility and capacity of the patient, as a consumer or user of health and disease-related care services, to weigh different alternatives with (approximately) the same indication and to select the most suitable according to their sense." Cf. Struppek 2010, 62.
302 Cf. http://www.weltagrarbericht.de/themen-des-weltagrarberichts/ernaehrungssouveraenitaet/ernaehrungssouveraenitaet-volltext.html [2017-10-17].
303 Cf. http://www.saatgutkampagne.org [2017-10-17].

contexts, from various uses of the term data sovereignty[304] to the notion of filter sovereignty, which refers to one being empowered to extract relevant information from a flood of data with the filters and analytical tools of one's choice.[305] On the other hand, this inflation in the variety of contexts and ways in which people assert claims to sovereignty contrasts with a loss of sovereignty experienced in real terms, measured, at least, by the classical meaning of sovereignty. Talk of digital sovereignty thus at times resembles as a wishful mantra or placatory ritual.

Cultural-historically, the concept of sovereignty is primarily rooted in the religious-political realm. While its evolution in the history of ideas has hardly been without major discontinuities, much can still be learned from its cultural-historical reconstruction. Sovereignty was regarded as that quality of a (monotheistic) God by virtue of which He is absolutely capable of doing or granting anything without reference to any other power or authority. In his early modern theory of the state, Jean Bodin adopted this claim to total independence and unconditionality to define and legitimise the sovereignty of absolutist rulers: their rule is limited by no other power, they are morally and legally accountable to no one, and they embody the highest authority in their domain (after God).[306] Such absolute and unconditional power always harbours the potential for capriciousness and abuse.

Presumably, such instances of absolute unconditionality and the potential for despotism are the reason that claims to

304 The concept of data sovereignty was used, for example, by politicians to describe an extension or renewal of data protection concepts, especially with a view to data economy (see https://www.heise.de/newsticker/meldung/IT-Gipfel-2016-Merkel-plaediert-fuer-Datensouveraenitaet-statt-Datenschutz-3490629.html [2017-10-17]). The Bertelsmann Stiftung uses the term to investigate how the "control" of its own data can be possible (cf. Bertelsmann Stiftung 2017) and the Bundesverband Digital Wirtschaft e. V. (German Association for the Digital Economy) used the term at a conference in June 2017 as a catchword for the current challenges presented by digitisation, without further defining its content (cf. http://www.data-summit.de [2017-10-17]).

305 Cf. Seemann 2011, 79.

306 Cf. Bodin 1976, 39 ff.

sovereignty have been subjected to critique, not only in political science and state theory, but also in ethical debates.[307] Against this background, concepts of sovereignty have been developed which, rather than emphasise the supposed absolute independence of the sovereign subject, instead point to the dependencies that are inherent to a subject's physical and social corporeality.[308] Such a notion of sovereignty, however, need not lead to a fatalism that simply resigns itself to the limitations and dependencies of one's personal way of life – on the contrary.

The further development of a moral-philosophical concept of sovereignty for the digital domain (see chapter 5) can profit from such nuances. While it is conceivable that a person could, in a state of absolutely unbound sovereignty, open themselves up to or close themselves off from any interest in accessing their personal data, this is unrealistic. Such an absolutely unbound sovereignty could, however, find a place in a morally coherent as well as realistic concept of digital sovereignty if interpreted from the perspective of the religiously and historically efficacious theme underlying all thinking about sovereignty. At least according to one reading, the point of this understanding of sovereignty – and thus its significance as a cultural-historical theme – is that the absolutely unconditional claim to sovereignty attributed to a monotheistic God in fact developed as a critical reaction to the perceived arrogance of the sovereignty claimed by worldly powers – who indeed all too often celebrated themselves as deities (e.g. Egyptian pharaohs or Roman emperors). Transferring absolute sovereignty to God thus ruled out any attempt by humans to establish sovereignty (in the sense of ultimate power of disposal) over other humans. The idea of divine sovereignty served, so to speak, as a regulative idea by which strict limits could be placed on powers of disposal amongst human subjects. This power-critical

307 Cf. Klein 2016.
308 Cf. Böhme 2008, 188.

motif to understandings of sovereignty can similarly counteract the temptation to insist on the digital sovereignty of each individual such that they come to possess an absolute power of disposal over data that is collected and processed in any way. If one follows this interpretation of the concept of sovereignty, it follows that, in principle, personal data is merely loaned to those who collect and use it; never is it at their disposal as property. Conversely, this means that data providers are not automatically the owners of their data, nor can they realise their claims to sovereignty under all circumstances (see chapter 3). Again, such a notion of sovereignty need not lead to a fatalism that resigns itself to the limitations and dependencies inherent to one's personal way of life.[309]

Of course, claims to sovereignty can be transferred to others, e.g., to representatives or, in the case of previously fully sovereign nation states, to supranational organisations (such as the European Union). Such transfers sometimes develop a momentum of their own, which can lead to considerable and practically irreversible losses of sovereignty on the part of the transferring party. To nonetheless insist upon the original sovereign's claim to sovereignty, even in the face of overwhelming evidence to the contrary, however, protects the former from total appropriation by the de facto sovereign and restricts the latter's exercise of sovereignty.

In principle, it follows from this understanding of sovereignty that the individual possesses extensive capacity for control, which at the same time is bound up in a variety of contexts in which it interacts with other agencies. In terms of an appropriate regulatory approach, this suggests a multi-actor perspective (see section 4.7). For example, while the "right to delete" certain kinds of personal, non-anonymised data may appear feasible at the individual level, in the context of mass data collection and processing, it can only be effectively exercised as a collective responsibility (see chapter 4), that is, as a

309 Cf. ibid., 194.

kind of right to inspect or monitor the work of data-processing institutions. Furthermore, it should be determined whether dynamic and cascade models of consent can be employed at least in part in the wider healthcare sector, in a way analogous to the procedures already in place at biobanks (see section 4.1.2). In the sense of the discussion in section 4.6 below – which concerns the issue of solidarity and an understanding of sovereignty that is also determined by the vulnerability of others – it must be asked whether sovereignty may not also encompass accepting the renunciation of certain means of control and of revoking consent. In addition, sovereignty is by no means limited to the idea of protection, but also includes a right of participation: namely the right to open access to the online world and, in particular, to the knowledge and new insights generated by big data.[310]

Sovereignty is realised through the exercise of power; conversely, it is limited by the exercise of sovereign power on the part of others. This reciprocal limitation of individual power does not mean that sovereign actors exercise mutual domination. This can be elucidated with the help of the functional distinction between "power *to* (something)" and "power *over* (someone)", which has become a common feature of recent theories of power.[311] In its universal-reciprocal function of limiting the power of others, the sovereignty of individuals is realised only as a power *to* something, namely to carry out legitimate, self-determined action, and not as power *over* others.[312]

The concept of power covers a broad spectrum of different meanings and refers to a barely comprehensible variety of phenomena. A definition that is both comprehensive and at the

310 Cf. Kettemann 2015.
311 Cf. Lovett 2007.
312 Of course, the illegitimate crossing of that border by third parties triggers defensive powers. In individual cases, these may even extend to self-help through self-defence and thus legitimise a specific form of exercising power over third parties.

same time meaningful is impossible.[313] However, the plausible starting point for a conceptual analysis of individual or personal power may still be the definition proposed by Thomas Hobbes, the first modern theorist of power, in "Leviathan": "The power of a man, to take it universally, is his present means to obtain some future apparent good".[314] This encompasses both conceptions of *power to* and *power over*, and is thus plausible as a general definition. In its radical abstraction, however, it is almost devoid of content. The same applies to Max Weber's sociological definition, which, in contrast to Hobbes, only captures the concept in the sense of power over others, i.e. as "the probability that one actor within a social relationship will be in a position to carry out his own will despite resistance, regardless of the basis on which this probability rests".[315]

In recent decades, more specific concepts of power have been proposed. These do not primarily take as their starting point the actor exercising power, but rather examine the social field in which power relations unfold,[316] on the one hand, and the objectives pursued or parties addressed by the exercise of power, on the other. The latter perspective is evidently concerned with forms of power over others. Beyond the use of violence and direct coercion, which do not play a role here, one can further distinguish between two modes of exercising such *power over*: firstly, the manipulation of the preferences and convictions of others (e.g. through the dissemination of

313 Power belongs to those complex concepts whose widely ramified web of meaning can hardly be forced into the bracket of a general definition (i.e. standardised in terms of necessary and sufficient conditions). Such concepts only become comprehensible in the mode of determining their "family similarity" with paradigmatic (and consensual) basic types of the term; cf. Wittgenstein's "Philosophische Untersuchungen", §§ 65, 66 (Rhees 1969).

314 Hobbes 1999, 66. The somewhat odd-looking inclusion of the word "apparent" designates the purely subjective determination of the desired good by the holder of power, which Hobbes declared decisive.

315 Weber 1978, 53.

316 For example, the concept of power as "power through control of the agenda" in Bachrach/Baratz 1962, which refers above all to a form of political power that manifests itself by limiting or influencing what is currently negotiable and in need of a solution. For the purposes of our analysis, this form of power does not play a significant role.

social myths), and, secondly, going beyond this influence, the manipulation of other subjects themselves[317] by subtly shaping, modifying, and thus ultimately potentially dominating their characters.

This is especially relevant here: the deployment of big-data algorithms at the interface between the providers and the users of online services enables the former, among other things, to exert strategic influence over the thoughts, feelings, actions and, ultimately, lives of the latter. Such abilities are the manifestations of new forms of interpersonal power. This alone does not constitute a compelling reason to deem them inadmissible without exception. However, some of these abilities are inadmissible and others may be subject to the need for special justification. In the context of big data and health, the key issue is not so much an obvious exercise of power that compels people to disclose their personal data against their will. While such coercion may exist as a legitimate legal constraint, it is then justified on the grounds of security, for example in certain areas of public health. Restrictions of sovereignty, however, are more often the result of more subtle, concealed power factors, i.e. the control of preferences that people unknowingly acquire and unthinkingly allow to influence the decisions they make.[318]

For the sake of clarity, some further distinctions are in order. The possibility of influencing others' convictions and dispositions covers a broad spectrum of conceivable forms. Not all of them appear morally dubious or specifically in need of justification. In this sense, at the far, ethically unproblematic end of the spectrum are various means of convincing others through the power of reasonable argument. In close proximity would be mildly paternalistic, nevertheless "libertarian" attempts to influence certain behaviours with the help of subtly

317 Cf. on the first aspect Lukes 2005, 25-37, on the second, Wartenberg 1990 and above all the numerous (unsystematic and fragmented) analyses by Michel Foucault, mainly in Foucault 1976. On Foucault's theory of power see Kneer 2012.

318 Cf. Lukes 2005.

motivating mental nudges. Well-known nudges in the domain of health and healthcare include redesigning cafeterias using principles of behavioural psychology in order to encourage people to choose healthier meals; insurers providing bonus points or rewards to policyholders who participate in voluntary sports programmes; and spatially and structurally redesigning offices and buildings to establish more movement and exercise in the routines of those using such facilities.[319] The transparency of such measures, their focus on users' welfare and their appropriateness to the matters they seek to address attest to their legitimacy.[320] On the other hand, forms of influence that employ suggestive strategies may be seen as morally suspect. This notwithstanding, they are also, in principle, socially acceptable. Passing over a few further points on our spectrum, we arrive at its other, sinister end, for example, at subliminal – i.e. not perceptible by their targets – interventions at the fundamental level of subjectivity – i.e. the ego – designed to manipulate actions, preferences, convictions and character traits for the benefit of other, self-interested parties.

All these forms of more or less subtle exercise of power by mentally invasive means are also conceivable in connection with the use of big-data mechanisms. For their normative assessment along a sliding scale of ethical value, a criterion for differentiating between admissible and inadmissible methods is needed. Deception seems to present itself as such a suitable criterion. Most forms of manipulation are likely to operate with a strategy of targeted misleading. However, by no means

319 For an overview and further examples, see Roberto/Kawachi 2015; Halpern 2015 or Buyx 2010.

320 The investigation and differentiation of such means of influencing third parties is the main subject of so-called behavioural economics and is therefore primarily pursued from an economic point of view. But nudges have also long been considered and applied as forms of legitimate political action. In its project MINDSPACE, the Institute for Government of the British Prime Minister has developed a model of modern governance based on six prerequisites. Without exception, they imply a self-determined participation on the part of citizens: "explore, enable, encourage, engage, exemplify, evaluate" (cf. Institute for Government 2010, 9). Cf. also Bröckling 2017.

does deception always imply manipulation. Simple lies, e.g. regarding one's circumstances at a given time, may deceive their addressee, but do not necessarily manipulate them. Conversely, manipulation is perfectly conceivable without deception, for example by means of direct intervention in the brain.[321] For the concept of manipulation, deception is therefore neither a necessary nor a sufficient condition. It is therefore not at all suitable as a criterion for the normative assessment of manipulation.

A suitable criterion would have to refer coherently to the object that the categorical definition of inadmissible manipulation is designed to protect. This can be nothing other than the self-determination of the person targeted by a manipulative intervention.[322] In this sense, self-determination refers not only to the concrete motive of a given action, but also to the entire field of possible motives a person may have for all their potential actions, to the cognitive as well as the emotional and motivational "environment" that shapes their actions, and thus to the "social space" of their existence as a social being.[323] Any manipulative change made to another person's social sphere of action translates to an exercise of power over that person and thus a potential impairment of their self-determination.

If self-determination is to be protected, it follows that the decisive criterion of inadmissibility is the manipulative exercise of power. The inadmissibility of manipulative interventions does not consist merely in the sheer weight of their consequences (however, this is to be determined).[324] Rather, they are inadmissible, or at least particularly in need of justification, to the degree that they circumvent the ability of their targets

321 Cf. Merkel et al. 2007.

322 Cf. again section 4.1.2 and the differentiations developed for the concept of self-determination (following Joel Feinberg).

323 See on this the terms "action-environment" and "social space" in Wartenberg 1990, 74, 85.

324 Even serious harm caused by third parties is legitimate in principle (up to the limit of life-threatening danger or permanent serious damage) if covered by the informed consent of the person concerned.

to control the conditions of their own actions and thus undermine or call into question their self-determination. This occurs when the manipulation *directly* targets the motivational field of an individual's potential action and thus escapes their cognitive control.[325]

The use of big-data algorithms at the interface between the providers and the users of online services and social media opens up a multitude of novel possibilities for exerting subliminal and thus inadmissible manipulative influence. A study carried out by Facebook on the principle of "emotional contagion" in January 2012, in which almost 700,000 unsuspecting Facebook users served as research objects, provides an illustrative example.[326] In this study, Facebook researchers pursued the (quite legitimate) goal of answering the question as to whether emotional contagion can be transmitted between people outside of the boundaries of highly personal communication, i.e. exclusively via digitally mediated contacts. According to the basic idea of the project, this could be determined by substantially reducing the quantity of certain kinds of expressly emotional content in Facebook users' news feeds, i.e. the reactions of other users to whom they are connected over the platform. The target group of 700,000 users was randomly divided into two subgroups: over the course of a week, 350,000 subjects received significantly fewer positive emotional reactions, while the other 350,000 received significantly fewer negative emotional reactions than they had actually been sent. Big data enabled the study's designers to filter out specific emotional language and conceal it from their subjects. Over the following days, also using big data, they then scanned the comments posted by these 700,000 unsuspecting participants, filtering out and classifying their specifically emotional content.

325 The expanding possibilities of such unauthorised, direct interventions into the "egos" of their targets are therefore increasingly countered by the human rights postulate of a "right to mental self-determination"; see Bublitz/Merkel 2014.

326 Cf. Kramer/Guillory/Hancock 2014.

The result: the 350,000 users who were undersupplied by their "followers" with positive emotions produced a significantly higher number of grim, cheerless or otherwise morose statements than the members of the group who had been "spared" others' negative emotions.

The authors of this study responded to the massive public and scientific criticism it prompted[327] by pointing to Facebook's data policy and terms and conditions, which every user must accept, and which expressly provide for the use of data in "studies".[328] Thus, they argued that affected users had already effectively consented to their participation. This is patently false: while this justification refers to the mining of user data for research purposes, which can be, depending on circumstances, wholly unproblematic, it completely ignores the manner in which this study intruded into users' mental spheres. Such intrusions are inadmissible in the absence of informed consent. Users' sensory or cognitive control over external interventions into the space of the ego were completely undermined. Therefore, this study was not simply about observing users by analysing their data – again, a potentially admissible action. Rather, at its core it was about effecting manifest changes to their mental state – a sustained manipulation of their emotional experience stretched over days, which, for half of these involuntary research subjects, was clearly for the worse. The fact that not one of the researchers may even have read a single comment or post submitted to Facebook by a participant during this study, and that their analysis was instead conducted entirely by machines, does nothing to change this.

In principle, similar attempts to influence the convictions or preferences of individuals or groups without their knowledge and consent – in such a way that transgresses the boundaries of what is widely accepted as legitimate manipulation

327 Abundant evidence on the scientific blog "The Laboratorium" by James Grimmelmann (cf. http://laboratorium.net/archive/2014/06/28/as_flies_to_wanton_boys [2017-10-17]).

328 Cf. https://de-de.facebook.com/full_data_use_policy [2017-10-17].

through advertising and the like, and instead problematically undermines control mechanisms essential to communication – would also be possible in the health and healthcare domain. Better known and far more widespread is consumer influence through the targeted tracking of online behaviour, in combination with other data pertaining to given individuals and groups (see section 2.4.1), such that tailored advertising can be delivered to them.[329] This sort of influence would be even more problematic if it affected subjects themselves, i.e. if it represented a data-supported, subtle shaping and modification of their character.

These examples illustrate well the possibilities, risks and, above all, normative limits of techniques based on big data that penetrate into the mental space of the data provider. If such interventions circumvent the individual's capacity for control, they are – absent special justifications (such as informed consent or in certain emergency situations) – in any case morally dubious or inadmissible, representing the exercise of a new form of illegitimate power. Of course, such subliminal interventions only become legally inadmissible if they lie outside the sphere of socially acceptable, generally tolerated behaviour. Most forms of psychologically sophisticated advertising that target people's unconscious motivational potential, or subliminal nudges for morally unproblematic purposes, therefore fall short of constituting violations of the law. Whether they are also morally acceptable, however, is a difficult question, one that is growing increasingly difficult as the possibilities for such psychological manipulation expand. Ultimately, this question must be answered for each individual case on the basis of the normative boundaries being developed here.

Upholding the concept of sovereignty, as it has been outlined here, and putting it into practice through various models

329 Cf. https://ca-commercial.com/casestudies/casestudyemployeebenefitsprovider [2017-10-17] as well as Datta/Tschantz/Datta 2015.

of consent tailored to the realities of big data (see section 4.1.2) does not simply mean that one can eliminate these uses and abuses of power, whether blatant or subtle. However, it does allow one to insist on the legitimate right to contain such power, so that the individual is not harmed by others exploiting the innumerable data points they leave in their wake.

4.4 Non-maleficence and beneficence

In the ethical analysis of big-data applications in the healthcare sector, another point of normative reference – one that is significant not only from an individual but also from a socio-ethical point of view – is provided by the moral duty to beneficence. One of the fundamental convictions of Western ethics is the assumption that, in many situations, moral action should extend beyond mere non-maleficence and produce a positive surplus particularly in the sense of improving the conditions of those in need. Depending on the cultural context, however, this elementary intuition has been concretised and shaped in very different ways. Within the Judeo-Christian tradition, for example, it has found its way – via the commandment of love for God and neighbour – into a comprehensive ethos, which is based on certain religious and metaphysical presuppositions and that therefore, despite its enormous effectiveness, cannot simply be generalised.

The call to beneficence took on a different form in modern theories of classical utilitarianism, which sought to define morally correct action by means of calculating its quantitative and/or qualitative utility. This of course entailed the danger of not granting enough moral weight to personal boundaries[330] and of sacrificing the fundamental rights of the individual to

330 While individuals as legal subjects are to be protected precisely in their individuality, classical utilitarianism tended to morally devalue the boundaries between individuals and to propagate collectivist objectives such as "the greatest benefit of the greatest number".

the supposedly greater good of an aggregate benefit for everyone affected by the consequences of an action.[331] Up to the present day, various teleological and deontological schools of thought have been developed in an effort to overcome this weakness inherent to classical utilitarianism.[332] Without wanting to reconstruct in detail the multifaceted development of the concept of beneficence, three systematic requirements can nonetheless be formulated, which appear indispensable to any ethically convincing recourse to this category.

Firstly, it must be made clear that recourse to a correctly understood concept of beneficence in the context of a medical course of action by no means amounts to an open or concealed utilitarian standpoint, with all the serious problems of action theory, epistemology and legal philosophy this would entail.[333] Secondly, the manner in which beneficence guides the normative idea of human goodness must be unpacked in such a way that, due to its abstract character, makes as little reference as possible to the particular ideological assumptions of both the agent and the recipient of action. This can be achieved, for example, by demonstrating the importance of a given beneficial act for the development of basic human abilities, without which it would seem inconceivable that the individual could thrive.[334] Thirdly, beneficence – characterised by its orienta-

331 Cf. Rawls 1972, 3-8.

332 See here Smart 1961; Smart/Williams 1973; Brandt 1992; Gesang 2003 as well as Gesang 2013 and Gesang 2000.

333 The concept of beneficence is by no means the same as that of benefit because of the former's inalienable attachment to individual rights; rather, the category of benefit that is central to utilitarianism represents a very specific, ethically highly controversial interpretation of beneficence that must not simply be identified with the benefit of a course of action (cf. Beauchamp/Childress 2001, 166 f.).

334 Such a (weakly essentialist) definition of the "good life", oriented to basic abilities, has the advantage that it is compatible with a multitude of ideas of happiness, as they can be found in modern liberal and pluralistic societies. From an ethical perspective, the opportunities and risks of technological innovations can only become the subject of normative directives if their respective relation to the harm or benefit of the pursuit of comprehensive human development for the individuals concerned is convincingly demonstrated.

tion primarily towards the welfare of the individual – must be concretised in accordance with rules of procedure and content in such a way that can avoid both an exaggerated altruism and a calculus that offsets the basic rights of the individual.[335]

Two aspects of beneficence should be of particular interest in connection with the topic of big data and health: the expansion of the knowledge base, on the one hand, and the improvement in therapies, on the other, that will result from new ways of extracting and processing information from massive volumes of data, and which stand to benefit a variety of participants. Even if, in modern societies built around information and (scientific) knowledge, it is considered self-evident that one would attach an intrinsic value to the continuous expansion of the knowledge base, this statement nonetheless requires an important elaboration, insofar as two perspectives need to be distinguished. On the level of the individual, it should first be established that knowledge and insight are of major importance for an individual's self-constitution and their ability to lead an autonomous life. Without an understanding of oneself and the world that, at least in its basic features, corresponds to reality, one can neither successfully form a personal identity, nor act autonomously or responsibly. Furthermore, the critical examination, verification, and expansion of knowledge also serves an important general, societal function. Functionally differentiated, complex societies characterised by the division of labour rely in their development upon stable, reciprocal expectations as regards the behaviour of others, which compensate for the increased contingency of social relations. The transparent communication of verified knowledge and the methodologically controlled expansion of bodies of knowledge are thus a key factor in ensuring reliability and social coherence.

335 Cf., for example, the list of "rules of beneficence" in Beauchamp/Childress 2001, 167.

The opposite dynamic is to be expected when trust in the reliability of others and their statements is diminished by deliberate deception (e.g. in the context of ideologically motivated propaganda, the targeted falsification of data or the dissemination of false or inaccurate reports), thus successively eroding social cohesion. To be able to correct such undesirable developments requires the preservation of communication that is committed to truthfulness, and, particularly in the sciences, sophisticated methodological and theoretical measures that have been developed to ensure this truthfulness. Thus, from an ethical perspective, it must not only be asked to what degree new digital techniques of data collection and analysis are associated with scientifically and philosophically relevant shifts in epistemic standards or losses in the reliability of the statements they generate (see section 2.2). Rather, it must also be clarified which groups of people will benefit primarily from the advances in knowledge that are actually brought about, how the current obstacles to a more efficient design of the data use process can be eliminated (e.g. through open-data and/or open-science strategies) and how a just distribution of the positive effects resulting from the expected growth in knowledge can be achieved (cf. section 4.5).

The generally held view that data is the most important resource of the 21st century may be especially true of the healthcare sector. Considering that, to date, high quality evidence-based standards are in place for only some instances and aspects of medical practice and that health-related decisions therefore must still frequently be made under conditions of high uncertainty, it is not surprising that the prospect of gaining insights and expanding the knowledge base through the targeted collection and interpretation of large amounts of data is one of the most important driving forces behind the dynamic development of health-related services.

However, not only must it be borne in mind that, as regards establishing and improving the knowledge base across the various fields of medicine and healthcare, the demands placed on

data quality and data security vary considerably. Expectations as regards the opportunities made available by such knowledge production also vary according to the actors in question: While basic medical research is primarily concerned with improving the understanding of certain fundamental, health-related processes, an attending physician is primarily interested in optimising the relationship between diagnostics and therapy that is relevant to patient care. The individual patient, on the other hand, will measure the benefit primarily in terms of the actual improvement in their well-being. Yet another perspective on potential opportunities may be found among those healthy people who collect and evaluate health-related information for the purpose of individual prevention. They may well do so with the view to benefit society as a whole, or a particular pool of insured, or at least out of super-individual concern, but they may also simply be pursuing their own economic ends. Since it is currently impossible to realistically arrive at an overall balance of the advantages that collecting and evaluating large amounts of health-related data will bring across the domain of health and healthcare, the opportunities and risks respective to each segment of medical service provision (e.g. diagnostics, prediction, therapy planning and implementation), context of action and affected group must be determined separately.

With the help of digital data processing, not only can early detection, diagnosis and therapy recommendations be significantly improved; new opportunities are also opened up for improving patients' quality of life. For example, the development of new sensor technologies makes it possible to adapt therapy more precisely to the needs of the individual patient. Since physician-patient contact is often hampered by spatial distance, or the necessary specialists are difficult to reach, the use of telemedicine, especially in rural areas, can contribute to more timely medical consultation and can lastingly improve the general quality of communication between physicians and patients without lengthy and costly physical visits to the doctor.

Even before a disease has manifested itself, the collection and evaluation of health-related data harbours considerable preventative potential. Not only can it help those with certain medical dispositions better avoid group-specific risks, but the monitoring of bodily functions and patterns of movement, or the optimisation of eating habits, with the help of specially designed wearable devices, can promote the pursuit of lifestyles conducive to good health. There is great optimism that the rapid developments being made especially in this area will provide a positive impulse for major improvements in healthcare.

These applications and those like them show that the potential of big data in the healthcare field is hardly limited to the ability to rapidly produce new knowledge. Rather, the complex concept of benefit has an economic and social dimension in addition to the purely epistemic, such that the collection and interpretation of data may yield completely different configurations of advantages depending on the type and motive of the actors involved: while individual citizens, for example, can collect certain kinds of personal health data primarily in order to optimise their behaviour for preventive purposes, the incomparably larger volumes of data pertaining to given patient cohorts primarily serve, within the framework of basic and clinical research, to improve the understanding of the emergence and development of pathological processes and to fruitfully apply the knowledge so gained to improving the diagnostic and therapeutic work of physicians. However, private firms have also become major players in the field of massive data collection, hoping to exploit the knowledge generated from this data for commercial product development – extending far beyond the pharmaceutical sector. Insurance companies are also likely to have an increasing interest in being able to better assess the risk profiles of their customers by compiling large amounts of health-related data and to influence their preventive behaviour through targeted financial incentives.

4.5 Justice

The collection and dissemination of large amounts of health-related data also touches on fundamental questions of justice, insofar as it concerns, on the one hand, the distribution of the anticipated opportunities and liabilities among various groups of people, and, on the other hand, the regulation of access to existing collections of data. Against the background of the overlap between publicly funded and private-sector data collection, both of these issues hold considerable potential for conflict.

Concepts of justice have always served to set the basic coordinates that morally and legally frame human relationships. Developments were already underway within ancient philosophy such that, rather than simply employing the category of justice to refer to a state of comprehensive virtue, internal differentiations were introduced, and various dimensions of justice were terminologically set off from one another. Thus, Aristotle distinguished so-called general justice (*iustitia generalis*), in the sense of action in conformity with the law, from various forms of special justice (*iustitia specialis*) such as distributive justice (*iustitia distributiva*) and compensatory justice (*iustitia commutativa*). The distributive justice which, according to Aristotle, regulates the relationship between the state and the various social groups that support the community and which takes as its criteria the respective contribution and/or dignity (*axia*) of these actors, opens up a space for the idea of what can be referred to as meritocratic justice. By contrast, compensatory justice refers to the right of citizens to exchange among themselves and regulates, on the one hand, the economic exchange relationships arising from the need for a sufficient supply of goods in terms of their ability to meet these needs, and, on the other hand, the need for compensation through legal avenues that may arise in cases of inadmissible maldistribution favouring certain actors.

Within this process of conceptual differentiation, two insights in particular have gone on to exert a major historical influence: Firstly, the idea that, because justice is a concept only grasped with reference to the other (*pros heteron*, *ad alterum*) – a fact recognised early on in the history of ideas – this not only renders unavoidable the need to differentiate between various social roles (e.g. as a citizen and legal subject or as a private contractual partner) and their corresponding types of relationships, but also makes it possible to extend the claim to justice across broad areas of social interaction. The modern idea, rooted in the 19th century, of comprehensive social justice as the epitome of a well-regulated social order, an idea in which all previously differentiated particular dimensions of justice are subsumed, only represents the end point of a long development that had powerful consequences for contemporary political awareness, a development that enshrined the principle of justice as the ultimate arbiter not only of individual action but also the basic institutional structure of a community.[336] On the other hand, it had already become clear in ancient discussions of justice that the basic, elementary demand of justice, to give to each their own (*suum cuique*), requires the elaboration of a differentiated system of criteria that takes into account not only the special nature of the good in question, but also the function and purpose of the particular relationship concerned.[337] In modern discussions of justice it is above all distributive justice that has undergone an important conceptual evolution that goes beyond the traditional contrast between needs-based conceptions and meritocratic conceptions of justice. While the idea long prevailed that the compensatory social effects demanded by the principle of justice could be readily achieved by means of financial transfers, representatives of the so-called capabilities approach[338] have rightly pointed out that

336 Cf. especially Rawls 1972.
337 Cf. Walzer 1983.
338 Cf. Sen 1979; Sen 2009; Daniels 1990; Nussbaum 1998; Nussbaum 2014 and Nussbaum 2015.

people in different social situations are by no means capable of making the same use of certain goods, and that the conceptual foundations of justice discourse thus need to be subjected to a renewed clarification. Against this background, the debate surrounding the current varieties of so-called participatory justice[339] and capabilities-based justice[340] is being constructively carried forward. Participatory aspects also play a central role in the current discussion of political justice. At stake here is no longer merely the question of the legitimacy of certain forms of rule in general, but also of the specific opportunities for access to and participation in political processes in which the rules of social interaction are negotiated and defined for everyone[341], especially for certain, particularly vulnerable groups.

As manifold as the concept of justice may be in terms of the history of ideas, the basic structure of its meaning can be determined as follows: as a normative principle of social relations, justice requires that any arbitrary privileges enjoyed by individuals or certain groups be overcome by rationally determining what is appropriate to each in each case, by giving equal consideration to this in the actions of others, and by requiring that any differences in its treatment be justified and that this justification is capable of finding consensus.

For the specific subject area of big-data applications in the healthcare sector, the following four problem areas are especially important from the perspective of the ethics of justice:

Firstly, the question of regulating the conditions of access to data collections is important for the research domain. At both the national and international levels, there exist various obstacles to the timely exchange and reciprocal availability of already established databases within the scientific community. Not only clinics and private companies, but also individual research teams tend to regard the data they collect as their

339 Cf. Bormann 2006.
340 Cf. Dabrock 2012.
341 Cf. Forst 1996 and Forst 2007.

property, even if public research funds or public infrastructure have been used to collect and analyse it. The withholding of data from the public sphere in this manner – which is problematic from the perspective of social and political justice – also occurs when the private providers of health apps hoard the data they have collected even though the development and use of the technology on which their services rely has been publicly funded (see chapter 2).

It is indisputable that – depending on the context – certain kinds of data should not be generally accessible to everyone, but rather must be protected from arbitrary access at the hands of third parties by whatever suitable means. Equally understandable is the interest on the part of actors in the private sector to generate profits from data collection and use. This commercial activity must, however, be balanced against legitimate interests arising from, where applicable, any previous reliance on public funding. It may be difficult or even impossible to apply the classical notion of private property, in the sense of an exclusive power of disposal, to health-related data that has been expressly collected for certain research purposes. Unwarranted barriers to access not only generate avoidable extra costs by compelling the repeat collection of data; they can also result in the potential of data already at hand not being optimally exploited, or in publicly funded data collection ultimately serving purely private, commercial ends. For the sake of fairness, it therefore appears necessary to employ suitable instruments in order to ensure that authorised persons have access to existing, research-relevant databases that is as unhindered and timely as possible. Such instruments could include open access or public science strategies, specific publication obligations, etc.

Secondly, there is a problem touching on issues of participatory justice wherever insidiously monopolistic structures establish themselves, rendering it difficult or even impossible for researchers to pursue their goals, or preventing individuals or groups from exercising sovereignty in determining how they will be affected by the use of their own data (see section

4.3). Challenges for participatory justice could also exist with regard to more subtle forms of power (those mentioned in section 4.3). This would be the case, for example, if data monopolies were used in a non-transparent manner in order to exclude individual users from certain services from the outset.

Thirdly, the rapidly growing market for health apps and various wearable devices used for private self-tracking also raises new questions of justice insofar as these products are predominantly aimed at supporting healthy lifestyles. As regards the preventive behaviour of individual health-conscious policyholders, in order to ensure an equitable structuring of health insurance premiums, it will thus be necessary to balance the idea of meritocratic justice that rewards healthy choices with the basic conditions of a healthcare system financed by and grounded in the principle of solidarity. The potential positive effects these developments could entail, both in terms of the improved health of those individuals concerned and in terms of cost reductions for the healthcare system as a whole, should not be denied, but rather emphasised. However, scepticism is warranted towards the, at times, accompanying idea of atomised individual responsibility that abstractly defines the fitness and ability of people or groups in total isolation from the social, economic and cultural contexts in which their lives are embedded.

Of course, active and conscious preventive behaviour always demands a certain element of personal ability in the form of commitment and individual effort, but it must not be overlooked that this ability always depends on certain conditions, some of which are beyond the control of the individual. Particularly if adopting such preventive measures is to be rewarded within the framework of certain health insurance premiums, it must be ensured, for reasons of equal opportunity, that everyone affected is given the opportunity to consciously decide for or against availing themselves of such offers as part of how they plan their lives, not merely in an abstract-theoretical sense, but in a real-practical sense.

At this point, considerations of justice and solidarity are intertwined; the latter will be dealt with in more detail in section 4.6. An overly narrow understanding of solidarity can give rise to discussions around the question of whether, in a system of care financed through solidarity, the entitlements of those who are unable on their own to meet the standards of preventive behaviour expected should be limited or even excluded. Real examples can already be found in the private health insurance sector. By contrast, the concept of justice, understood both in terms of the capabilities approach and in terms of equal opportunity, as well as an inclusive understanding of solidarity, demand that the use of health data be limited if the basic solidarity-based structure of access and financing within at least the statutory health insurance system is to be preserved (see section 4.6).

Lastly, the themes of social inclusion and an appropriate consideration of the manifold social conditions that come to bear on the individual exercise of freedom have found their way into an integrative notion of capabilities-based justice, which generally serves to protect all the goods and rights proper to the individual and constitutive of the conditions on which the personal capacity to act depends. In an increasingly digitised environment, these conditions also include the – certainly demanding – ability to deal responsibly with one's own (and other's) health-related data. Depending on the context of action and the functional role of the respective actor, various aspects have to be considered: while as thorough a clarification as possible of the type, scope and purpose of data collection with regard to the design of diagnosis and therapy seems necessary from the patient's point of view, as a precondition of truly informed consent, someone privately using digital services in order to optimise their own health and lifestyle choices may be primarily interested in actively shaping the process of sharing their data and in being able to realistically assess the quality of the data interpretation provided. Even an approach oriented towards the ideal of capabilities-based justice does not

demand that the individual affected is able to perfectly understand all the technical details of modern information gathering and processing systems. Nonetheless, the constitutive ability to arrive at informed judgements – an indispensable qualification for providing valid consent for, or issuing a justified prohibition against, the collection, evaluation, storage and dissemination of one's data – must be strengthened and fostered. In this sense, the concept of a capabilities-based justice, or a justice of capability, refers to an understanding of data sovereignty that will be explored in more detail in chapter 5 below.

4.6 Solidarity

The historical development of the concept of solidarity is characterised by heterogeneity, and today the term is understood very differently across various disciplines; as a result, there exists a whole series of parallel ideas about solidarity.[342] However, some core elements can be identified that are characteristic of most understandings of solidarity. Accordingly, solidarity refers, roughly speaking, to prosocial actions, practices and dispositions, as well as institutional, political and contractual arrangements, which are intended to "lend support to others, or at least [...] express the inclination to want to help and support".[343] Most authors agree that such prosocial inclinations or attitudes must also be accompanied by certain appropriate actions or the bearing of certain costs[344] in order to be able to speak of solidarity.

Solidarity is usually seen as being complementary – and often subsidiary – to justice. This applies in both an ideal and an institutional sense. While understandings of justice are subject

342 For a detailed overview of the genesis of the term solidarity and its various permutations, see Prainsack/Buyx 2017.
343 Bayertz 1996, 308.
344 "Costs" are understood here in a broad sense (including financial, social, emotional, temporal, etc.).

to numerous rules codified under the rule of law, duties pertaining solidarity are usually assigned to the level of social morality. The various understandings of justice outlined above decidedly refer to everyone; at the institutional level, they are underpinned by the neutrality of the state in passing judgement on the ways individuals plan their lives, choose their values and objectives, etc. It is precisely these concepts and the valuations associated with them, however, on which the practice of solidarity and its corresponding duties and responsibilities come to be based. The underlying motivation for lending support to and bearing costs for others in actions of solidarity is usually based on the "recognition of relevant commonalities with another person or persons".[345] Such commonalities can consist, for example, in the perception of dangers, existentially significant risks, etc., which are beyond the ability of any one person to defend themselves from alone. Solidarity thus arises from a common goal pursued by a group, either in the face of a common challenge or threat, or arising from a shared idea of the good life within a community of solidarity; this idea is expressed, for example, in the form of the German welfare state. A community based on solidarity shares essential views of what offers of mutual of help should be made available to its members and why; its members are therefore prepared to make contributions in order to realise a common vision.

Solidarity is thus not neutral, but is always oriented towards a specific, substantial goal or a shared understanding of how a community should be structured. Put another way, it could be described as standing between the just and the good. Solidarity is also not always automatically good in the normative sense; such a categorisation depends on how the commonalities and objectives that motivate and generate solidarity are to

345 Prainsack/Buyx 2016, 82. Because of this recognition of commonalities, relationships of solidarity are symmetrical in relevant respects and, especially when institutionalised, are characterised by an element of mostly indirect reciprocity (ibid.).

be judged.[346] Solidarity is particular; its content and goals must be determined anew in each context. This applies to concrete practices of solidarity within circumscribed social groups such as neighbourhoods, associations, self-help groups, etc., as well as to the large community of solidarity that constitutes the system of statutory health insurance (Section 1 SGB V). The objectives guiding solidarity, especially at the higher levels of institutionalisation, can be quite abstract: for the statutory health insurance system, for example, it is "the task of maintaining or restoring or improving the health of the insured" (ibid.). The community of solidarity within the statutory health insurance system establishes – via the principle of need, on the one hand, and the principle of individual financial capacity, on the other – that risks of disease and injury are to be borne jointly by all members, i.e. by all insured. This could be described as an act of solidarity by those who are, at a given time, stronger on behalf of those who are weaker, regardless of the specifics of their cases.

As the logic of social insurance thus illustrates, the readiness to act in solidarity often depends on expectations of reciprocity. This does not mean that people who put solidarity into practice tie the support they lend to others directly to a payment or benefit they will receive in return, for this would not be solidarity at all, but simply the business of reciprocal exchange. The relationship between solidarity and reciprocity is more fundamental: it consists in the expectation that one will be the beneficiary of the solidarity of others if requiring support in the event of an analogous case of injury (illness, etc.). This is the principle of health insurance, of statutory accident insurance, but also of insurance providing for property loss or damage, travel cancellation or the need for legal aid.

346 There can be found, for example, numerous examples of structural solidarity groups in criminal organisations, e.g. in the mafia or in terrorist organisations.

It is nonetheless inadmissible to equate solidarity with reciprocity. If help and support are extended solely or primarily in anticipation of a (concrete) return payment of whatever sort, this, again, would not be an act of solidarity. However, in practice, solidarity – and especially institutionalised and sanctioned solidarity – are bound up with elements of reciprocity in an important respect. Wherever the rules of solidarity are legally codified, wherever duties are made statutory, as is the case with health insurance, (indirect) solidarity plays a crucial stabilising role. For all the common commitment to the necessity of healthcare for all, the costs of such an inclusive and impersonal structure of solidarity would not be borne by its members without the guarantee that everyone who contributes can also expect help in the event that they themselves experience injury or loss. The willingness to act in solidarity can certainly decrease if the impression arises that others' need for help and support is a result of their own negligence or lack of initiative, thus over-stretching the fabric of solidarity.

Even this brief description shows that communities of solidarity, especially those that institutionalise statutory or contractual obligations, are in constant need of justification, both in terms of their objectives and their membership.[347] While in Germany the fundamental mutual social security function of the statutory health insurance system is rarely called into question, the determination and demarcation of membership in this system, as well as the nature and scope of the benefits and services funded by solidarity premiums, are certainly the subject of controversy. It has long been debated whether the exclusion of sections of the population by virtue of their membership in the private health insurance system violates the fundamental principle of solidarity, namely that it pertains to society as a whole. Currently, the main topic of discussion is whether and how full membership status in the statutory health insurance system can be extended to new sections of the

347 Already discussed by Rorty 1989, 189-191.

population. Communities of solidarity thus have a certain tendency to include only those who are already part of them, and thus to exclude others, whom one may or must refuse (certain) claims to solidarity.

4.6.1 Solidarity in statutory and private health insurance

The question thus arises as to whether and to what extent big data in the healthcare sector produces, or could produce, changes to the existing pattern of solidarity in the statutory health insurance system and beyond. As outlined in chapter 2 above, the analysis of lifestyle and health-related data opens up a wide range of new opportunities for risk stratification. Big-data processing can produce much more accurate risk profiles of people. This represents the continuation of a trend that began with clinical risk scores (the classic example being cardiovascular risk factors) and which has been the subject of controversy since the emergence of the first precise genetic tests (especially regarding monogenetic diseases such as Huntington's disease[348]), at the latest. Even then, it was feared that, given the new possibility of identifying groups, in an evidence-based manner, with low or high (or, in the case of monogenetic autosomal diseases, certain) risks of illness, some of these groups, namely low-risk groups, might want to leave the community of solidarity or withhold their financial contributions. In any case, it was expected that others, namely high-risk groups, would be the targets of discrimination.[349] However, because of the presence of numerous other disease risks

348 Huntington's disease (also called Huntington's chorea) is a very rare autosomal dominant heritable neurodegenerative disease. It is caused by a gene mutation in which the nucleic acid sequence cytosine-adenine-guanine is repeated more than 36 times on the fourth chromosome. Typical symptoms are movement disorders and changes in character to the point of dementia.

349 Cf. Chadwick/Berg 2001.

that cannot be revealed using these methods, as well as largely effective anti-discrimination laws, these fears have not, or have only rarely, been proven justified.[350]

By contrast, big data offers so many further possibilities of risk differentiation that these concerns are increasingly moving back to the fore, especially as regards the preventive capacities and behaviours of individuals, which, by means of data-rich methods, can be tracked with ever increasing precision, as mentioned above. If all data pertaining to a person, from genome sequencing, microbiome and proteome data, to other health and lifestyle information, to leisure and shopping behaviour, social contacts and browser use, can be analysed in an integrative manner, the justified assumption is that risk profiles could be generated of a completely different order of sophistication and, more importantly, with a completely different, much higher predictive power.

The basis of the community of solidarity within the system of statutory health insurance is the shared vulnerability of *all* members to disease risks that cannot be anticipated and quantified with certainty. For this reason, all members are subject to the obligation to jointly assume responsibility – by collectively financing a statutory fund – for the treatment costs incurred in all cases of illness suffered by individual members. If, in the future, disease risks – with the exception of accident risks (and some infectious diseases) – can be determined as precisely, multidimensionally and reliably as appears to be the case, then this basis could be called into question by a kind of fragmenting logic of exclusive solidarities.

Globally, there already exist a number of examples of increased stratification amongst insured, while in Germany structural changes to the insurance system resulting from the new possibilities of risk stratification are the topic of regular

350 A well-known precedent are the 2004 court proceedings concerning the lifetime appointment to the civil service of a teacher affected by Huntington's chorea, who finally won her appointment in court. See the judgment of the Darmstadt Administrative Court in NVwZ-RR 2006, 566.

discussion.[351] In Germany, strict limits are placed on the use of such stratification procedures for excluding people from benefits within the statutory health insurance system, but this is not always the case elsewhere. Even the basic constitution of the German healthcare system – a structure and community of solidarity – is being acutely challenged by current developments.

Within the dual structure of the German health insurance system, in which statutory health insurance and private health insurance coexist, the precision and refinement with which potential policyholders can be screened and selected with the help of big data poses a fundamental danger: private insurers could increasingly filter out individuals with favourable/low risk profiles, at the cost of the statutory system, which would thus be more heavily burdened with high-risk policyholders. This is likely to spur policy debates on the legitimacy of private health insurance in this context. In any case, considerable controversy surrounds the question of whether the increasing function of the statutory health insurance system as a subsidised sink for the "bad" risks shed by private insurers calls into question the entire, underlying structure of solidarity within a national health system. Conversely, there is also the view that this stratification of risks makes a comprehensive statutory health insurance system more attractive.

Challenges to solidarity arise not only from the relationship between statutory health insurance and private health insurance, but also within the private insurance system itself. Private health insurance operates with risk premiums. Before a policy is taken out, comprehensive data is collected on a prospective policyholder in order to assign them to a premium group. This includes requesting medical records. If data on, for example, one's fitness regimen and dietary habits could also be integrated into this process, this could help generate even more accurate risk profiles – at least, this is the fear. Another

351 Cf. Andelfinger 2016; Deutscher Bundestag 2014; Radic et al. 2016; Heuvel 2016.

possibility would be the individual adjustment of premiums not only on the basis of developments within the premium group as a whole, but also on the basis of the continuous collection of personal data following the activation of the policy. This would entail the abandonment of the basic principle of insurance, whereby risks are to be borne collectively by a larger group of policyholders and premiums are not to be individually adjusted. It could also lead to the creation of increasingly small premium groups, in which claims would lead more quickly to premium increases.

In 2014, for example, the private insurer Generali Group announced that, together with the South African financial services provider Discovery, it would unveil in Europe a new product called Vitality, "a behaviour-based insurance model designed to motivate and reward customers for improving their health and well-being".[352] According to its press release, Vitality is based on three principles: rewarding health-conscious behaviour, regular contact with the insurance company and tailor-made programmes to improve lifestyle choices, and an innovative relationship between insurer and customer. However, this press release does not mention that the regular contact and tailor-made programmes are tied to the collection of customer data and the requirement that customers make this data available. Data pertaining to fitness activities and dietary habits comes from wearable devices or health apps running on a customer's smartphone or smartwatch; the reward offered the customer is, among other things, a premium bonus.

This raises the question as to what happens when a policyholder is unwilling to take part in a behaviour-based insurance model. The most reasonable assumption is that they would be withheld the privileges of lower premiums, bonuses, gifts

352 http://presseservice.pressrelations.de/pressemitteilung/generali-group--partnerschaft-zwischen-generali-und-discovery-zur-einfuehrung-des-innovativen-produkts-vitality-in-europa-581548.html [2017-09-19].

and vouchers, which over the long term would translate into a disadvantageous overall premium vis-à-vis their fellow policyholders. Regardless of whether their actual behaviour and lifestyle are conducive to good health, they would be punished for not making their data available to their insurer. Their exercising of their right to informational self-determination would thus lead to a concrete disadvantage. Furthermore, such insurance models would also withhold advantages from those customers who, as a result of sheer misfortune, are already at a disadvantage and unable to exhibit a favourable risk profile.

Finally, the ramifications that insurance models based on behavioural data would have for the statutory health insurance system are cause for particular concern. The solidarity principle operative in the latter is characterised by the fact that policyholders pay an income-adjusted premium regardless of risk and receive, in accordance with SGB V, the benefits they require. Data-supported risk profile generation, with the exception of some financial bonuses for preventive behaviour, is largely alien to the process of setting premiums in the system of statutory health insurance. It contradicts the principle of solidarity that, as described above, demands that protection be provided against vulnerability relating to disease or injury without regard to individual (behavioural) risks. The intensive discussion surrounding the fact that people, through wilful disregard or at least negligence, can be culpable for a disease or injury from which they suffer (e.g. ski accidents, nicotine-related illnesses, etc.) has revealed how complex the assignment of responsibility for certain health conditions can be at the level of the individual (the key term being "personal responsibility in healthcare"[353]).

Health is impacted by many factors that lie beyond individual agency or control. Even today, it is difficult to demonstrate a clear instance of personal fault that could be qualified as a violation of the expectation of indirect reciprocity within a

353 Cf. Langanke et al. 2013.

solidarity-based system and which could thus justify the withholding of benefits or other sanctions.

Moreover, according to the solidarity principle of the statutory health insurance system, the freedom to determine the course and conduct of one's life and develop one's personality takes precedence over a strict and permanent obligation to avoid all health risks. Of course, there are reasonable limits in this regard, as expressed in the passage on the responsibility of the insured for their own health in Section 1 of the SGB V. Nonetheless, it would be difficult to qualify acquiescence to the continuous, targeted collection of one's personal data and the use of risk profiles, fed by big data, which encompass all areas of one's life as a reasonable demand to make of someone in terms of their responsibility for their own health. Therefore, the question must be raised as to whether the generation of big-data-driven, highly predictive risk profiles for the purpose of classifying patients should be outright prohibited.

A longstanding debate has thus intensified considerably in the face of the new possibilities opened up by big data, a debate surrounding the personal responsibility of the insured and the question of whether and how this can be accounted for in developing the health insurance system under the unavoidable constraints imposed by scarce resources. One aspect of this debate is the question of whether it should be a central task of the statutory health insurance system to influence the health-related behaviour of its policyholders and to thus, as it were, provide public health education. The 2015 Präventionsgesetz (Prevention Act) calls upon the statutory health insurance funds to establish "quality-assured services for the promotion of health-conscious behaviour" (new version of Section 65a SGB V). According to the consumer advice centre of the state of North Rhine-Westphalia, the focus of bonus programmes within the system of statutory health insurance, which have been in place since 2003, should thus extend "from a customer loyalty scheme to an instrument for steering the behaviour of

policyholders".[354] There already exist numerous examples of this, such as collecting bonus points, which can then be redeemed in a rewards shop. Bonuses are awarded, for example, for proof of healthy physical activity, but also in some cases for simple pledges such as declaring that one is not a smoker. Thus, insurers do in a certain sense deliberately influence their policyholders' lifestyles, using rewards to guide them towards certain standards. In light of this, it is worth asking which behaviours and characteristics are assigned value, what this value is, and why certain behavioural patterns are considered worth encouraging. A longstanding aspect of the classic debate surrounding personal responsibility in healthcare is the fact that it is both practical and, from a normative perspective, very problematic to classify and compare certain behaviours or behavioural patterns in such a way, e.g. to place a value on physical exercise, because of its health benefits, vis-à-vis something like involvement in one's community. Data-derived incentive structures could develop into highly intensive and invasive forms of surveillance. Nonetheless, this only addresses one side – the negative side – of the developments resulting from the increasing use of big data in the healthcare context.

4.6.2 New practices of solidarity

To conclude, some possible positive developments with the potential to reinforce solidarity should be highlighted. 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 – and that the occurrence of serious illness is a certainty only amongst very few groups of people, namely

354 Verbraucherzentrale Nordrhein-Westfalen 2015, 18.

those with genetically determined, especially monogenetic diseases. To exclude or place at a disadvantage such groups would be inadmissible on well-known, ethically sound and legally guaranteed anti-discrimination grounds.

However, other phenomena that are already present and worthy of consideration – and potentially of being fostered – can also be associated with the concept of solidarity. In various areas of medicine, the application of big-data technologies has already led to the development of new, prosocial mutual support practices. This indicates that the use of big data in the healthcare sector not only has the potential to erode and undermine solidarity as described above, but also to improve and strengthen solidarity in other ways. Much as in the discussion surrounding questions of self-determination, whether such gains can be realised depends upon the context in question and the technologies and intentions involved.

Examples of new practices of solidarity include, for example, the formation of small groups of patients who share the same risks or even concrete experiences of – especially 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. Social scientific studies of such phenomena show that participants explicitly point to shared experience and the motivation to help others with similar experiences that this engenders.[355] Such initiatives often emerge within the framework of solidarity-based support groups, e.g. patient self-help groups, which are becoming increasingly popular in light of the expanding possibilities for digital data storage amongst laypersons. 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,

355 Various references in Prainsack/Buyx 2017.

the expansion of such practices is to be expected.[356] Whether they represent sustainable gains for patients or again harbour the risk of developing an exclusionary dynamic as a result of the increasing sophistication and predictive power of the data on which they are fed, remains unclear. In any case, this would depend on various factors, such as the motives behind data sharing and the quality of data preparation. From the perspective of solidarity, there are at the moment good reasons to encourage the further development of these practices, but in a way that is transparent and that pays heed to the possible dangers involved.

4.7 Responsibility

In light of the multiple options for action and decision-making at the individual, collective, state and non-state levels, as well as the many institutional and legal regulatory and design possibilities that exist in this context, it must be asked who is the acting subject to whom these options are available and to what degree. Traditionally, an ethical inquiry along these lines is associated with the concept of responsibility. There exists a broad consensus that responsibility is a moral category for the orientation and judgement of action and inaction.[357] It can be differentiated according to types of action and decision, but also according to the design of institutional structures.

The scope and degree of responsibility that can be ascribed a subject depend upon their internal and external capacities, professional competence, power, and spatial, temporal or

356 An example of this phenomenon is the online platform iManageCancer, which is being developed by an EU H2020 project (cf. http://imanagecancer.eu [2017-10-17]). In addition to various apps for health information, there is also the possibility of data storage and networking both between patients and with treatment algorithms; the aim is for patients to better live with cancer and to themselves take responsibility for some aspects of their treatment – including outpatient chemotherapy (dose adjustment, treatment of side effects, etc.) – in consultation with their physicians.

357 Cf. Heidbrink/Langbehn/Loh 2017.

social proximity to the occurrence that calls for an ascription of responsibility. These differentiating criteria form a basis for the attempt at answering the general question of responsibility – who is responsible for what, with what aim, why, to whom or what, when, and how –; i.e. the question of subject, object, purpose, reason, instance, time, and the other circumstances pertaining to responsibility.

The answers to these questions may be found in various spheres. Responsibility can be both demanded and assumed morally, legally, politically and contractually, and for all of these spheres, this applies both retrospectively and prospectively, i.e. before and after an action or decision takes place. 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. Responsibility can exist for both actions and omissions. The reasons for and consequences of this ethically and legally significant distinction are the subject of considerable controversy. Debate also surrounds the question as to whether – in complex, structured, highly technical work or decision-making processes – collectives, organisations, institutions or even machines, especially self-learning machines, can be held responsible alongside the individuals involved. The complex interaction between people (as individuals, but also as the performers of various roles and functions), institutions and technology is of particular importance in the context of the use of big data in the healthcare sector and thus deserves special attention. 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.

4.7.1 Individual responsibility as regards the sharing of health-related data in various roles and contexts

It is one of the commonplaces of classical data protection law and its underlying moral principles that the collection, storage, processing and disclosure of personal data is permitted after informed consent has been provided and with due regard to other data protection principles (acceptable uses of data and the obligation to adhere to these, the necessity, proportionality and minimisation of data collection, and a ban subject to the possibility of authorisation; see chapter 3). In the English-speaking world, "notice and consent" has become the established term for this approach. It states that after at least formally acknowledging the conditions of data collection and processing, and thus indicating that one can reasonably be assumed to be informed, the individual who has provided their data is essentially responsible for the consequences of the ensuing process. The technical complexity involved in describing the possible consequences – as formulated in a typical list of terms and conditions, a legal prerequisite for contractual validity – is, however, likely to overwhelm the typical user of data-intensive processes or devices. Because this typical user normally has the choice of either assenting to terms or conditions that are difficult to understand and whose consequences are difficult to foresee or refraining altogether from using the service or device in question, for most people there is no real alternative. It is thus often impossible to speak here of responsibility in an objective and substantial, rather than merely a formal, sense, at least according to a definition of responsibility that takes as its prerequisites a sufficient degree of personal capacity to act and a sufficient number of alternative courses of action.

Nonetheless, criticism of the notice and consent approach, as it is commonly understood, does not obviate the need to assign to the individual certain responsibilities and to expect of

them certain, corresponding actions and attitudes, even in the age of big data. The care that ideally should be taken by users, but which is often strained by verbose and opaque terms and conditions, should thus be directed towards observing long-standing advice: e.g. handle personal data with care; closely scrutinise "free" services and offers; pay attention to the location of data processing as well as any privacy settings; encrypt any data that is to be used by multiple vendors; set up a secure wireless network; be cautious when using smartphones that collect and share considerable amounts of data (location and contact information); keep software up to date; turn off devices regularly, etc.[358] These maxims are, however, most easily followed in the big-data era insofar as a framework of conditions is in place for the individual that technically and organisationally enables easy and effective use.

Furthermore, if the use of big-data applications is based on a certain social role that is bound to a specific ethos in which others can trust – for example, the role of a doctor or a scientist – there exists a special responsibility that implies increased due diligence. Physicians have the duty to ensure that the software used in their practice, any practice network to which they may belong, or the transfer of patient data for payment purposes, etc. function securely and are not used for purposes for which no assent has been granted. Precisely because of the impossibility of reliable data anonymisation discussed above, scientists and researchers for their part are expected to take special care with regard to the accuracy of data collection, processing and transmission and, where appropriate, the programming of algorithms.

4.7.2 The responsibility of institutional actors

The complexity of big-data processes, hardly comprehensible for laypersons and often difficult even for experts to

358 Cf. Schaar 2014, 265-276.

understand, means that the companies and institutions involved have special moral and legal responsibilities. These companies and institutions are the parties often most likely to fulfil the requirements of a plausible attribution of responsibility: they possess the skills, competences and power to take whatever steps necessary to exercise control – or refrain from taking steps that undermine control – and to guarantee both with appropriate mechanisms. It is irrelevant whether, in terms of the inner workings of an organisation that is processing data (public authority, company, research institute), this moral responsibility can be legally attached to a person; a moral responsibility exists because of the aforementioned fact of this organisation's abilities in its external relationships.

Among the ways that companies can responsibly design big-data processes, the most important is to create the conditions that allow data providers to revoke their consent. Of course, a data provider is in theory always granted the ability to bar the continued or future use of their data by refusing to participate, by not selecting certain options, or by simply deleting their account. In practice, however, the data provider's position would be significantly strengthened if they could effect and guarantee the retroactive deletion of their data. It is both technically and morally conceivable that data could be administered on demand. Sufficiently aggregated data, derived data or models that demonstrably do not allow conclusions to be drawn about data providing individuals could be exempted from deletion. To use such approaches in order to enable the decontextualisation and recontextualisation of data while ensuring high standards of anonymisation and creating trust in institutions is likely to be one of the key tasks facing future medical research.

Another way for industry to take responsibility for the rights of the individual while also protecting legitimate business interests would be to set up delegation systems at the programmatic interfaces in data networks, through which most data exchange takes place (see chapter 2). Since these data

networks usually do not feature interfaces that can be used directly by individual data providers, software systems – i.e. algorithmic actors – who, as *data agents*, could represent people and their interests in the world of machines, may represent the most practical solution. Such software tools could include predefined rules for the handling of data that have been developed by third parties (consumer organisations, etc.) and would interact automatically with the systems that companies and institutions use to collect and process data.

Thus, the generally unreasonable task of data providers themselves manually managing their data in detail is replaced with a programmatic system of management. The advantage of this model consists in relieving the data provider of the need to develop special tools in order to manage their own data. The operative principle is to provide the individual with representation in the form of a software system operating in their interest; it provides the individual with the technically low-threshold ability to take responsibility for choosing their own short-, medium- and long-term strategy for the handling of their data, without having to decide every single question themselves.

It would also be helpful for the individual if organisations and entities that collect, process and disseminate data made it technically possible to shape the incessant data flow in ways that afford more transparency for the individual at those points that concern them. However, it must first be clarified what is meant by transparency and how it can be created. Calls for the publication of all algorithms seem to entail exaggerated expectations, considering their complexity and dynamism. Such a complete form of disclosure would also complicate, or even render impossible, the protection of trade secrets, which are indispensable to any business seeking to operate in a market economy. To be sure, companies could be obliged to make their procedures verifiable and thus more accountable. One can technically verify, for example, whether a model or algorithm achieves a desired or claimed level accuracy. It is only

legitimate to demand and ensure that procedures do not systematically disadvantage certain population groups. Auditing at a system level could ensure that rules on data retention, anonymisation or deletion are followed. In many cases, such a review could be carried out by external, independent bodies or left to the companies themselves in the form of an indirect, process-oriented review.

Transparency is time and again demanded of the organisations that collect, process and pass on data so that individuals are given the opportunity to monitor not only their own datasets, but also the transfer of this data to third parties. To meet this ethically legitimate demand requires a technical infrastructure that protocols the origin, processing, use and exchange of data in a seamless and tamper-proof manner (see section 2.2) and which enables the data provider to form a clear picture of how their data is being used. In the healthcare sector, the question of how such organisations, with the help of technical systems and transparency procedures, can meet their responsibilities – ensuring that data providers remain as sovereign as possible within data networks, ensuring that they retain agency and are not hopelessly entangled in these systems – is of particular importance because the data concerned is highly sensitive from the outset. Clinical practitioners, manufacturers of medical devices, but also insurers, credit institutions or credit agencies must thus be aware of their responsibility to maintain the highest standards of data quality and the best data security procedures. For with each data input, self-learning algorithms will continue to rewrite themselves in ways that are often not comprehensible in detail.

In order to maintain the quality of data and processes at the highest possible level, state-level regulations may need to be considered. Another possibility, which has by now been widely considered, would be strengthening confidence in the organisations and processes concerned by means of certification, quality seals, or voluntary standards established and overseen by interest or industry groups. The extent to which this could

work can be read from comparable stories of success and failure across various industries.[359] If it were possible to establish a seal of quality that is as uniform as possible and which also shows data providers that their needs are being efficiently taken into account, the institutions concerned will be moving in a responsible direction. This would require, for example, verifying and responding to complaints in a timely manner, providing equally rapid corrections or making available alternative channels for lodging complaints. This would proactively enable the individual to achieve and defend their data sovereignty, a goal that will be formulated in the following chapter.

Besides the questions of access to and the transparency and use of data, there is an additional issue of responsibility facing organisations that collect, process and pass on data: whether, when, and to what extent they are in certain situations prepared to intervene in the personal communication that they enable amongst their users and which they – to the extent that this is permitted by the terms and conditions of use – also analyse. In 2017, the news that Facebook had developed an algorithm with which it hoped to be able to identify and reach out to people with suicidal intentions garnered considerable interest.[360] Evidently, this is a situation in which different maxims of legal and ethical responsibility collide. On the one hand, the inadmissibility of blatant and unsolicited intrusions into the private, or indeed intimate, sphere of someone who is dealing with a problem that is best solved by talking to one's closest friends and family or by seeking professional help speaks against this practice. This kind of intervention on the part of the social network also prima facie reeks of paternalism. However, if the functional reliability of such an algorithm were scientifically well substantiated, one would have to take account from an ethical perspective of the fact that in this case

359 Cf. https://www.splendid-research.com/Marktforschung/Guetesiegel-in-Deutschland-2013.pdf [2017-10-27].
360 Cf. e.g. http://www.bbc.com/news/technology-39126027 [2017-10-17].

one is literally dealing with a question of life or death. Suicide is the second most common cause of death amongst youth in Germany and young people in particular use social networks intensively.[361]

While in the case of suicide risk, weighing these values that exist in a state of tension with one another may yield a balance in favour of offering or even being obliged to offer help, in the case of less dramatic health issues this calculus becomes more complicated: should one send unsolicited offers of help – e.g. advice to visit a certain doctor – to members of social networks or identifiable users of search engines whose activities online create the impression that they may be at risk of disease? Does the answer to this question vary according to the severity of the disease suggested by a user's data footprint? Undoubtedly – while bearing in mind the issues concerning the consent model in the age of big data outlined in section 4.1.2 – one would tend to be on safe footing if one had obtained a data provider's consent for these potential health-related interventions in advance. There will certainly exist grey areas with regard to finding solutions to such problems, for example, when an indication of a potentially worsening health problem may exist, which can currently be addressed without too much effort or significant risks of health-related stigma. Here, too, the responsibility of institutionally organised actors is an important factor.

4.7.3 The responsibility of state bodies

Confronted with issues like the intersection of big data and the healthcare sector – a field characterised by numerous actors with different potential responsibilities – one is likely to quickly call for some form of state intervention. It goes without saying that we are not dealing with a legal vacuum. Nonetheless,

361 Cf. the interview with Jakob Henschel in "Der Spiegel" of 11 March 2017: https://magazin.spiegel.de/SP/2017/11/149997421/index.html [2017-10-17].

because of the rapid technical developments described above, various legal regulations that had long proven adequate are no longer as effective as they were at the time of their enactment. It is also true that the unsuitability of many provisions results especially from the fact that, in a globalised world, many national and EU regulations are difficult to enforce and easily circumvented. It is thus necessary to ask how the state can assume responsibility at the national level, within the EU, but also as an actor within the context of international law. The key analysis of this question has been outlined in the legal section above (chapter 3). It remains here to point out some legal-ethical fundamentals for the orientation of state responsibility. In light of the aforementioned problems of enforcement, a principle of regulatory subsidiarity should apply: wherever self-regulation and certification procedures are able to effectively secure the desired outcome, detailed legal regulations should not be issued, at least until the proven emergence of a significant problem that cannot be solved in this manner. Limiting state intervention to the guarantee or verification of a framework for action should always take priority over the effort to issue a comprehensive system of detailed regulations. The goal should be to connect legal clarity, practicability and flexibility to each other as coherently and productively as possible.

Quality seals, which would be established at the initiative of organisations that collect, process and pass on data or, failing this, with support from state authorities, could also help individuals handle their data in a sovereign manner. This would particularly apply if such seals were – in the ideal case of a legally enforceable regulation – respected by all or – in the case of a voluntary commitment by industry – as many providers of apps, services, studies, etc. as possible. It should also be explored whether it may be possible to offer the individual user more than the simple alternative of granting or withholding consent to the entirety of whatever they are offered, by instead making available a differentiated set of options concerning maximum storage time, data economy, deletion, data

migration etc. Similarly, data providers should be given the information they need to be able to at least roughly weigh the potential benefits and risks of whatever individual offer they avail themselves of within a given service.

4.7.4 Conclusion: multi-actor responsibility

If one considers the three levels at which responsibility could be assigned in the field of health-related big-data applications (individuals, organisations, government), it is those organisations that collect, process, and pass on data that, measured by the decisive criterion of their technical ability to introduce, implement and establish measures, will be seen as having the primary obligations. These organisations should ensure the conditions for the responsible shaping of informational freedom on the part of the data provider. Of course, the primary responsibility of such organisations does not exempt individuals from also assuming responsibility, as data providers, for the use of their data. However, it does not suffice to simply trust in the ability of users to rationally interact with the terms and conditions they encounter in the digital world or to believe that by carefully and sparingly availing themselves of digital applications, users can sufficiently guard against unwanted uses of their data. What is needed instead is support from those actors that collect, process and pass on data. The less these 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. It should be clear that 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.[362]

362 Cf. Braun/Dabrock 2016.

5 DATA SOVEREIGNTY AS THE SHAPING OF INFORMATIONAL FREEDOM

Existing big-data applications as well as those that are emerging entail both enormous opportunities and risks that must be taken seriously. This applies with particular force to the healthcare sector, as outlined above in chapter 2. Hitherto unimaginable possibilities are opening up in research, in new approaches to diagnosis and therapy in medicine and public health, but also in new ways to shape personal health on the part of the individual. Alongside these developments, and in connection with them, novel and economically highly attractive business areas are emerging. However, because the opportunities presented by big data can only be realised when data is collected, analysed and linked on a massive scale, these opportunities are also associated with new risks, especially threats to freedom, privacy, sovereignty, security and well-being, as well as challenges to societal solidarity, social justice and the necessarily assuming of responsibility by a number of actors that interact with each other in complex ways (see section 4.7).

Among the key characteristics of big data is the fact that certain kinds of data can no longer be assigned a priori or inherent degrees of sensitivity. Because of recontextualisation and recombination, new and potentially highly problematic statements can be derived from data that arises from contexts of use that appear, in themselves, neutral. In view of these conceivable consequences, it should be obvious that particular attention and care are called for, especially in connection with the processing of health-related and medically relevant data. While it cannot be ruled out that in some contexts (e.g. clinical research) a growing willingness to share certain kinds of data can yield positive effects and is thus to be conditionally encouraged, there also exist other contexts and fields of action in which an exceedingly liberal handling of personal data can

entail grave consequences and therefore must be subject to the highest standards of protection.

Below, the term data sovereignty is used to refer to the core aim of doing equal justice to the opportunities and risks posed by big data and of accepting and mastering the actual and normative challenges with which they are associated.

5.1 Data sovereignty as a guiding concept

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 growing role of big data.

The notion of shaping informational freedom builds on the concept of informational self-determination, while developing it further. It is not grounded in exclusive rights analogous to property rights, but rather in each person's authority to determine with which content they choose to relate to the wider world and thus how they wish to express and develop themselves communicatively. Shaping informational freedom in this sense refers to the interactive development of one's personality in a networked world while at the same time preserving one's privacy. Privacy (see section 4.2) ought not to be understood here in too static and rigid a sense as a (permanently) closed off space, but must rather be grasped in a manner that is flexible and context-dependent and structured accordingly. The shaping of informational freedom understood in this way is characterised by the capacity to intervene effectively in the constant flow of individually relevant data on the basis of personal preferences. The shaping of such freedom is responsible when it also orients itself towards the legal and societal demands of solidarity and justice.

The specific notion of data sovereignty as the responsible shaping of informational freedom must be distinguished from other possible interpretations. With the concept of data

sovereignty as described here, we seek neither to perpetuate the established, barely modified regulatory approach of data protection, only under a new name, 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 German Ethics Council seeks 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.

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 unsupervised, free development of their personality in the public sphere – as long as this does not necessarily enter into conflict with the rights of others. Everyone must be able to determine themselves how and to whom they wish grant access to their person via the disclosure of their personal data. With the concept of data sovereignty, however, we also wish to emphasise 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. Data must not be solely understood as an important, individual possession, but must rather be grasped in its collective dimension. The individual remains the determining point of reference for data sovereignty; beyond this, however, one must also take account of the relevance of data as a social resource. The guiding concept of data sovereignty developed here should therefore guarantee a high level of protection for the individual while at the same time making it possible to take advantage of opportunities at the collective level, for example in clinical practice and health-related research.

5.2 Data sovereignty in the healthcare sector

Data sovereignty in the sense outlined above is obviously not limited to the area of classical healthcare. On the contrary, the differences between the challenges faced there and those faced in other sectors are not of kind but of degree. Nonetheless, certain peculiarities specific to the handling of medical and health-related data can be highlighted: 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.

New technologies for collecting, analysing, and linking large amounts of data are driving improvements in our understanding of the emergence and development of complex health impairments, knowledge which over the medium to long term could be fruitfully exploited for innovative treatment concepts. Patients in clinical care can directly benefit from such concepts in two ways. For one, a more precise stratification of patients into medically relevant subgroups opens up new opportunities for sparing patients unnecessary stress resulting from unsuccessful procedures, in line with the medical directive to minimise harm. Furthermore, in line with the medical ethics principle of beneficence, big data promises to enable improvements in diagnostics, prediction and the planning of therapy; this could prove to be of enormous value to patients even if a complete understanding of many multifactorial diseases is still a long way off. The potential held by big-data applications to help people design and pursue health-conscious lifestyles also opens up new dimensions of prevention, which not only benefits the individual but also the community of solidarity that finances the health insurance system.

Nonetheless, these opportunities are paralleled by certain risks that, among other things, concern important aspects of

solidarity and justice. For the individual, there are risks of data protection violations and informational self-endangerment, which could lead, among other things, to discrimination, stigmatisation and thus to negative consequences in terms of fair participation and empowerment. Intensified risk stratification driven by big data could lead to an insidious or, if reflected in the premium structures of private health insurance, for example, explicit erosion of solidarity in the health insurance system and society more broadly that would impact especially vulnerable groups. The scientific community in turn could be challenged, for example, by an excessive increase in data of dubious quality or by problems in data exchange and interconnection. The risks posed by big data should not be ignored. They demand constant, critical consideration and should be minimised wherever possible.

Both individuals and organisations that work with big data must thus be empowered to act responsibly, in constant awareness of important risks, to take advantage of the opportunities presented by big data. At the same time, structural prerequisites, safeguards and protective mechanisms must be established in such a way that they allow the individual maximum freedom of action while minimising risk to the greatest extent possible. Particularly when one looks at the already complex German healthcare system and the diverse, often globally networked research sector and data industry, the importance of a differentiated, multi-actor perspective is obvious in order to effectively, efficiently, fairly and responsibly implement data sovereignty across various health and medicine-related contexts of action.

Against this background, two increasingly proximate and, in places, already overlapping spheres can be distinguished in terms of shaping and exercising data sovereignty: 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.

5.3 Outline of a governance concept oriented towards data sovereignty

The concept of data sovereignty is thus endowed with a specific character in the context of health and healthcare. It takes into account all of the considerations made above, namely the analysis of the scientific, technological and societal developments thus far (see chapter 2), the discussion of the legal framework (see chapter 3), and the examination of the fundamental ethical and normative conditions (see chapter 4), all pertaining to big data. At the same time, it constitutes the key criterion for the following considerations regarding a governance approach that, taken as a whole, can ensure a balanced and appropriate treatment of the opportunities and risks associated with big data and which will lead into the recommendations made in chapter 6.

The starting point for a governance concept that orients itself towards data sovereignty is a recognition of the futility of seeking to respond to big data with a blockade or interdiction. Clearly, developments in the field of big data that have already taken place – and which will only intensify in the future – cannot be stopped, but they can certainly be steered. Nonetheless, as the above legal discussion has revealed, the legitimate concerns that arise in this regard cannot be adequately addressed by the forms of actions and safeguards of traditional data protection law. To ensure that the constitutionally guaranteed right to informational self-determination can be truly and effectively exercised, the classical approach to data protection employs directives such as data economy or direct earmarking.

However, these instruments clearly reach the limit of their effectiveness in the face of big data. In the era of big data, it is neither possible nor sensible to apply them consistently. Practically no additional protection would result: demands along these lines are thus misleading. Meanwhile, the innovative potential of big data would be severely restricted or lost.

Even bearing in mind the criticisms that have been directed at the concept of informational self-determination – in the specific form it has assumed through decisions at the level of the constitutional courts and their implementation at the level of non-constitutional law, as well as against the background of the new GDPR – it follows from the fundamental normative considerations developed in this Opinion that sufficient legal and extralegal safeguards are needed to effectively counteract the dangers of power asymmetries and resulting losses of data sovereignty under the conditions of big data, particularly in the sensitive healthcare sector. In order to continue to be able to provide an adequate level of protection for the individual, the task ahead is to devise new models for regulating and shaping these developments that more faithfully reflect their complex and dynamic nature. These should reflect the principle of data sovereignty as the shaping of informational freedom in a multidimensional way and with a view to various groups of actors and contexts of action, taking up the potential forms and attributions of responsibility described above (see section 4.7).

Under the conditions of big data, such a multidimensional and multi-actor perspective must abandon outdated notions of a particular kind of data having a specific, given sensitivity that invokes 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. From a regulatory point of view, it follows from this insight that one should no longer focus exclusively or predominantly on the early stage of data collection and the notion of a degree of data sensibility that is largely context-independent and which can

already be established at this stage. Rather, greater attention must be paid to the various contexts of data use. Thus, the focus should be on the relevant processes and actors, and their respective responsibilities, at each stage of data collection, analysis and linkage.

A model designed to regulate and shape these developments in the era of big data that orients itself toward the principle of data sovereignty must focus on data providers 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 a sovereign relationship with their data, by designing regulations and shaping institutions in a manner sensitive to context and appropriate to each case. In light of the range of contexts in which data can be collected, analysed and linked in the field of healthcare, the diversity of these contexts and the degree to which they directly impact upon questions of self-determination and personality, this is an especially relevant goal. Cascading consent models (see section 4.1.2), which extend to the option of delegated or representative consent, hold particular promise in fulfilling the normative expectation of responsibly shaping the tension – implicit in the concept of data sovereignty – between the opportunities and risks posed by big data. The implementation of such models can be made more effective by additional technical and regulatory measures, such as the opening up of programmatic interfaces for data providers, data trustee models or mechanisms designed to improve the traceability of data and how it is being used.

Nonetheless, enabling and shaping data sovereignty is by no means the sole responsibility of those affected; rather, alongside basic agreement regarding a set of key normative parameters that must be secured through effective regulation, it requires a diversified analysis of the various spheres of responsibility and capabilities concerned and the areas in which they pertain. The shaping of data sovereignty cannot be achieved through the partial modification of certain instruments or by simply allocating additional resources or

expanding enforcement capacity, e.g. by upgrading and better equipping data protection officers. An increase in financial, material and human resources is undoubtedly necessary, but cannot be considered sufficient, at least within the framework of existing regulatory approaches. Rather, the task ahead calls for a comprehensive effort on the part of society as a whole – an effort that presumably can only be made globally – that would incorporate legal as well as non-legal mechanisms while also taking technical developments into account.

To this end, simplified, wholesale solutions should be avoided in favour of more complex, institutionally diversified, compound models that are more flexible and appropriate to the problems posed. In particular, what is needed is a pre-emptive, pre-structuring normative support that takes into account the technical design of instruments that intensively consume and process data. Overall, such a normative support would need to be geared towards ensuring a high level of confidentiality in data processing, especially in areas that belong to clinical practice and medically-related research or are closely related to them. In this respect, the two-sphere concept introduced above as regards the health-related domain can be further developed. The heterogeneous second sphere 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 there. According to the principle of regulatory subsidiarity and in order to avoid excessive regulation, however, incentives, voluntary commitments, certifications and other softer solutions should be relied upon to the greatest extent possible rather than hard, legal requirements backed by the threat of sanction. It is also important to encourage among the companies involved the realisation that, at least in the medium to long term, such a concept also offers competitive advantages vis-à-vis models that prioritise short-term commercial interests over data sovereignty.

To conclude, it is a matter of attuning all actors to the constitutive importance of the sovereignty of the data provider in every (health-related) data process, and to build upon and maintain this awareness going forward. It is to be expected that, in light of the likely unabated increase in the precision of sensor, data storage and data transfer technology, the two spheres described above will continue to draw closer to one another. This must not be allowed to lead to a reduction in the higher standards of quality and protection that prevail in the first sphere, i.e. in the domain of medical research and clinical practice. Instead, conversely, it is necessary to point to these standards and to insist that they can be upheld in the second sphere as well, that, when it comes to the handling of data, standards of quality and protection are not to be loosened but rather strengthened. Similarly, it is necessary to proceed according to this principle against any efforts to circumvent existing national and EU legal restrictions by shifting operations to regions and legal systems that are less sensitive to the need for data protection. Such intentions contradict the idea of data sovereignty advocated here. Towards the goal of implementing a governance in a responsible, effective and efficient manner, the German Ethics Council hereby issues a series of recommendations for action, taking account of the differentiation of spheres and the various permutations of responsibility outlined above.

6 RECOMMENDATIONS

In view of the manifold challenges posed by rapid developments in the collection, recombination and analysis of large volumes of data in the healthcare sector, the German Ethics Council sees the need for a fundamental change in the ways health-relevant data are handled and understood. In place of the existing model, which assigns to certain types of data static degrees of sensitivity, the constantly expanding possibilities of data decontextualisation and recontextualisation demand that the fluid, context-dependent sensitivity of data be emphasised.

Data sovereignty as a guiding concept

The German Ethics Council recommends a governance concept that orients itself towards the central goal of data sovereignty. We understand data sovereignty to be the responsible regulation and promotion of informational freedom, in a manner appropriate to the risks and opportunities presented by big data. To ensure this, traditional data-protection law – which primarily concerns itself with the constitutionally protected right to informational self-determination – must be further developed and redesigned in order to incorporate fundamental, normative guidelines, comprehensive in their content, while also taking new instrumental approaches. In order to take advantage of the opportunities presented by the use of big data in the healthcare sector, while also effectively countering the risks posed by new forms of asymmetrical power relations and the losses of individual sovereignty they could bring about, as well as the potential for discrimination, adequate and suitable protective mechanisms and design strategies are required. When traditional instruments – such as the prevailing strict focus on data minimisation and the narrow adherence to an intended and explicit data use – reveal themselves to be dysfunctional, other means of securing individual freedom and privacy and shaping a just society must come to the fore. The German

Ethics Council calls for a sophisticated, innovation-oriented governance concept that involves a large number of actors, paying attention to both the differences between and the increasing convergence and overlapping of two spheres: medical research and clinical practice, on the one hand, and the heterogeneous health-related products and services available on the free market, on the other. Such a concept calls for a comprehensive societal effort that includes both legal and extralegal elements, incorporates technical advancements and ensures their availability to all societal actors in such a way that guarantees the preservation of their basic rights.

The governance concept recommended by the German Ethics Council should encompass the following measures and ensure their timely realisation and funding:

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 Bundesdatenschutzgesetz [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 European General Data Protection Regulation (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 reidentification:

- 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 reidentification, 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, policy-making 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 a way analogous to medical ethics – which focuses on the benefit of the individual and, in accordance with the principle *nihil nocere*, holds as its highest maxim the avoidance of harm in every instance – the inalienable rights of the individual and their degree of self-determination must also be held as the ultimate measure of social progress when dealing with the opportunities and risks posed by the use of large volumes of data. These rights are not contrary to the common good, rather they are constitutive of a liberal and social state under the rule of law. By contrast, the need on the part of the (health) industry to acquire an ever-more penetrating insight into the expressions and manifestations of people's lives is not. On the other hand, big data entails great potential. For example, it can identify correlations between health and its social and societal determinants, and new ways to shape the most diverse areas of life in a manner that promotes health can be developed.

However, big data will prove itself beneficial for healthcare and medicine only if the individual, as the sole owner of their personal data, is empowered to decide at any time to whom they wish to disclose their data, and to which extent, even in cases of secondary use.

Data protection therefore demands precise legal regulation, and the Bundesdatenschutzgesetz (Federal Data Protection Act) itself requires specification with regard to suitable protective mechanisms and design strategies, i.e. a ratification and expansion encompassing data minimisation and earmarking. These functional instruments would ensure a consolidation of personal privacy and data protection and thus assist the implementation of the best possible form of data sovereignty. Data sovereignty must also maintain a higher status vis-à-vis research interests. Thus, a preliminary analysis of the possible ramifications of any new research procedure in terms of data protection and informational self-determination should

be carried out as a matter of necessity (i.e. a data-protection impact assessment). Data protection and the data sovereignty that comes with it – which, contrary to recommendation B1.3 above, also entails the ownership of personal data and thus an absolute power of exclusion vis-à-vis third parties – is an object of very considerable value and must therefore be secured by means of regulation and legal sanction. Only in this way can the state of self-determination which makes informed consent possible be guaranteed.

Technological advances in the ability to evaluate large volumes of data must be accompanied by legal restraints, such that the application of this technology is permitted, but abuse of the personal sphere is prevented. Analogous those contained in the Gendiagnostikgesetz (Genetic Diagnosis Act), strict prohibitions on the discriminatory use of personal data must be implemented. The storage and analysis of personal data should therefore only be permitted within a narrowly defined framework. Misuse of data, including secondary use, must be sanctioned by criminal law. The financially effective punishment of violations, capable of functioning as a deterrent, is key. It should be noted that this area is not characterised by a regulatory deficit as much as a massive enforcement deficit.

In this sense, the question as to whether big data offers opportunities or risks in the healthcare sector will be decided on the basis of the following conditions:

No central storage of patient data

In principle, patient data should not be stored on central servers, but on decentralised storage media in the possession of the patient. Additionally, encrypted patient data can be stored at the family doctor's practice, and to protect the interests of the patient, the decision to share this data (even to a limited extent) should be, wherever possible, made together with the family doctor.

In contrast to recommendation A4.1 above, any personal data, including those collected with the electronic health card,

must therefore not be stored centrally and not without strict individual consent in the case of primary and secondary use. The consent model that in principle requires close earmarking for the collection of any personal data must be retained. The introduction of a cascade model, which according to recommendation B2 above would include a broad scope of consent to data use, must therefore be rejected.

To ensure maximum data sovereignty and data security, it is imperative that small and medium-sized local companies, and the associations of statutory health insurance companies, implement these measures.

Consent on the part of the insured takes priority over others, as well as over research interests

In this sense, third-party access to decentralised data storage without the consent of the insured must be prohibited even for secondary use. The exchange and integration of health-relevant data across a variety of institutional actors recommended in A2 above must also comply with this principle. Technological safeguards must be met, such as effective anonymisation and pseudo-anonymisation measures that render reidentification impossible, which go beyond recommendation D.1.1.

Legal regulations in place of voluntary self-monitoring

Imposing and enforcing safeguards in this domain must be a matter of legal obligation and not voluntary self-monitoring and co-regulatory measures, as suggested in recommendation D3 above. Self-regulation does not work; in practice, the process consists of industry groups developing their own codes and creating their own (pseudo-)procedures with which to react to complaints. Self-regulatory bodies are given the responsibility of ensuring that their respective members abide by the rules and that remedial measures and sanctions are applied as needed.

To rely on co-regulatory measures, such as a system of industry-internal monitoring that presides over the awarding of

a "quality label", is naive. For example, the recent data-breach scandals affecting the data service provider Equifax in the summer of 2017, with 143 million affected customers, reveal the failure of internal control mechanisms.[363] According to warnings issued by the FBI, patient data is ten times more expensive than credit card numbers on the black market.[364] In 2015 alone, the health data of 80 million Americans were seized by hackers from Anthem Inc., the second-largest healthcare provider in the United States.[365] A clear framework for the handling of personal data must be created on both the regulatory and judicial levels in order to clarify what is allowed and what constitutes misuse.

On the technical level, data protection requires that security levels are already defined in the development phase as regards hardware, software and algorithms, and that they are appropriate to the type of data being protected. The right to provide consent, as well as "the right to be forgotten", must be afforded those affected, allowing them to track their personal data and, if necessary, to delete these data in any case and without exception. Exceptions, such as those proposed in recommendation B1.2 are not acceptable. Correspondingly, manufacturers must be obliged to ensure, early in the process of planning and manufacturing new products, that the basic principles of data sovereignty/data protection are adhered to without exception (privacy by design) such that complete privacy protection is ensured. Testing innovative concepts in clinical settings must be ensured primarily by public research.

363 See https://investor.equifax.com/news-and-events/news/2017/09-07-2017-213000628 [2017-09-11].

364 See https://www.reuters.com/article/us-cybersecurity-hospitals/your-medical-record-is-worth-more-to-hackers-than-your-credit-card-idUSKCN0HJ21I20140924 [2017-09-11].

365 See https://www.forbes.com/sites/gregorymcneal/2015/02/04/massive-data-breach-at-health-insurer-anthem-reveals-social-security-numbers-and-more [2017-09-11].

Conditions for donating data for the benefit of the public interest

Contrary to recommendation A4.2 above, creating legal options for individuals to donate their data for medical research, as well as making personal data available, can only proceed alongside strict and effective data protection with concrete earmarking. This would be a necessary condition for the benefit of the public interest and would complement the requirement that patient data be kept primarily on decentralised storage media in the possession of the patient. The Hasso Plattner Institute's health platform "Gesundheitscloud", for example, aims to provide patients with this kind of effective control over their own health data, as well as the option to altruistically donate effectively anonymised data for research purposes.[366]

Ownership of data

The individual's right to be informed about stored data pertaining to their person is no longer sufficient. Only when there exists full transparency concerning which data are fed into which analytical and evaluative processes, which criteria are used to classify these data, and how they influence decisions, can the legality and ethical acceptability of data use be determined. Contrary to recommendation B1.3 above, full "ownership" of data, which excludes third parties, must be ensured.

To conclude: If comprehensive data protection, the implementation of effective standards of anonymisation and pseudo-anonymisation, and the right to forget cannot be guaranteed, the use of big data for research purposes or other applications would, as a necessary consequence, have to be forgone.

Christiane Fischer

366 See https://hpi.de/open-campus/hpi-initiativen/gesundheitscloud.html [2017-09-11].

REFERENCES

Acar, G. et al. (2014): The web never forgets. Persistent tracking mechanisms in the wild. In: Association for Computing Machinery (ed.): Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. Scottsdale, 674-689.

Acquisiti, A.; Brandimarte, L.; Loewenstein, G. (2015): Privacy and human behavior in the age of information. In: Science, 347 (6221), 509-514.

Albrecht, U.-V. (ed.) (2016): Chancen und Risiken von Gesundheits-Apps (CHARISMHA). http://www.digibib.tu-bs.de/?docid=00060000 [2017-10-17].

Amma, N. G. B. (2016): Big data mining. In: M. K. Singh; K. G. Dileep (ed.): Effective Big Data Management and Opportunities for Implementation. Hershey, 53-59.

Andelfinger, V. P. (2016): Gesundheitsportale und Private Krankenversicherung – was in anderen Ländern passt, das passt auch in Deutschland – oder? In: Andelfinger, A.P.; Hänisch, T. (ed.): eHealth. Wie Smartphones, Apps und Wearables die Gesundheitsversorgung verändern werden. Wiesbaden, 105-107.

Augsberg, I. (2015): Artikel 8 GRC. Schutz personenbezogener Daten. In: H. v. d. Groeben; J. Schwarze; A. Hatje (ed.): Europäisches Unionsrecht (7. Aufl.). Baden-Baden, 142-158.

Augsberg, S. (2016): Big Data im Recht der Transplantationsmedizin – Vom „Ende der Theorie" zum „Ende der Aporie"? In: Medizinrecht, 34 (9), 699-705.

Aumann, I.; Frank, M.; Pramann, O. (2016): Kapitel 12 – Gesundheits-Apps in der Gesetzlichen und Privaten Krankenversicherung. In: U.-V. Albrecht (ed.): Chancen und Risiken von Gesundheits-Apps (CHARISMHA). http://www.digibib.tu-bs.de/?docid=00060000 [2017-10-17], 244-280.

Axer, P. (2011): Normengeflecht und Wissensrezeption in der gesetzlichen Krankenversicherung. In: Gesundheit und Pflege, 1 (6), 201-209.

Bachrach, P.; Baratz, M. S. (1962): Two faces of power. In: American Political Science Review, 56 (4), 947-952.

Barocas, S.; Selbst, A. D. (2016): Big data's disparate impact. In: California Law Review, 104 (3), 671-732.

Baumgartner, U.; Gausling, T. (2017): Datenschutz durch Technikgestaltung und datenschutzfreundliche Voreinstellungen. Was Unternehmen jetzt nach der DS-GVO beachten müssen. In: Zeitschrift für Datenschutz, 7 (7), 308-313.

Bayertz, K. (1996): Staat und Solidarität. In: K. Bayertz (ed.): Politik und Ethik. Stuttgart, 305-330.

Beauchamp, T. L.; Childress, J. F. (2001): Principles of Biomedical Ethics (5th ed.). New York.

Becker, M. (2017): Ein Recht auf datenerhebungsfreie Produkte. In: Juristenzeitung, 72 (4), 170-181.

Bedi, G. (2015): Automated analysis of free speech predicts psychosis onset in high-risk youths. In: npj Schizophrenia, 1, art. no.: 15030. DOI: 10.1038/npjschz.2015.30.

Bedi, G. et al. (2014): A window into the intoxicated mind? Speech as an index of psychoactive drug effects. In: Neuropsychopharmacology, 39 (10), 2340-2348.

Bennett, J. (2017): Your Face Belongs to Us. http://www.thedailybeast.com/how-facebook-fights-to-stop-laws-on-facial-recognition [2017-10-17].

Berdin, J. (2017): Biobank-Governance. Unter besonderer Berücksichtigung von Trust-Modellen, Baden-Baden.

Berlin, I. (1969): Four Essays on Liberty. Oxford.

Berndt, C. (2013): Resilienz. Das Geheimnis der psychischen Widerstandskraft. München.

Bertelsmann Stiftung (ed.) (2017): Rethinking Privacy Self-Management and Data Sovereignty in the Age of Big Data. https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/rethinking-privacy-self-management-and-data-sovereignty-in-the-age-of-big-data [2017-10-17].

Binder, J.-H. (2012): Regulierungsinstrumente und Regulierungsstrategien im Kapitalgesellschaftsrecht. Tübingen.

Bleicher, A. (2017): Demystifying the black box that is AI. https://www.scientificamerican.com/article/demystifying-the-black-box-that-is-ai [2017-11-17].

Bocanegra, C. L. S. et al. (2017): HealthRecSys: a semantic content-based recommender system to complement health videos. In: BMC Medical Informatics & Decision Making, 17 (1), 63.

Bodin, J. (1976): Über den Staat. Munich.

Böhme, G. (2008): Ethik leiblicher Existenz. Frankfurt am Main.

Bormann, F. J. (2006): Soziale Gerechtigkeit zwischen Fairness und Partizipation. John Rawls und die katholische Soziallehre. Freiburg im Breisgau.

Brandt, R. B. (1992): Morality, Utilitarianism and Rights. Cambridge.

Braun, M.; Dabrock, P. (2016): Ethische Herausforderungen einer sogenannten Big-Data basierten Medizin. In: Zeitschrift für medizinische Ethik, 62 (4), 313-329.

Britz, G. (2010): Informationelle Selbstbestimmung zwischen rechtswissenschaftlicher Grundsatzkritik und Beharren des Bundesverfassungsgerichts. In: W. Hoffmann-Riem (ed.): Offene Rechtswissenschaft. Tübingen, 561-596.

Bröckling, U. (2017): Gute Hirten führen sanft. Über Menschenregierungskünste. Berlin.

Bublitz, J. C.; Merkel, R. (2014): Crimes against minds: on mental manipulations, harms and a human right to mental self-determination. In: Criminal Law and Philosophy, 8 (1), 51-77.

Buchner, B. (2017): Art. 22. Automatisierte Entscheidungen im Einzelfall einschließlich Profiling. In: J. Kühling; B. Buchner (ed.): Datenschutz-Grundverordnung. Munich, 471-481.

Buchner, B. (2006): Informationelle Selbstbestimmung im Privatrecht. Tübingen.

Buchner, B.; Kühling, J. (2017): Art. 7. Bedingungen für die Einwilligung. In: J. Kühling; B. Buchner (ed.): Datenschutz-Grundverordnung. Munich, 260-280.

Budin-Ljøsne, I. et al. (2017): Dynamic Consent: a potential solution to some of the challenges of modern biomedical research. In: BMC Medical Ethics, 18 (1), 4.

Bundesverband Digitale Wirtschaft (ed.) (2017): Digitale Gesundheit. Die fünf großen Blockaden auf dem Weg zur digitalen Gesundheit – und wie wir sie überwinden können. Düsseldorf.

Buyx, A. (2010): Können, sollen, müssen? Public Health-Politik und libertärer Paternalismus. In: Ethik in der Medizin, 22 (3), 221-234.

Cartledge, C. (2017): Big Data. Data Wrangling Boot Camp. Big Data Vs. http://www.cs.odu.edu/~ccartled/Teaching/2017-Spring/DataWrangling/Presentations/030-bigDataVs.pdf [2017-10-17].

Caulfield, T. (2007): Biobanks and blanket consent: the proper place of the public good and public perception rationales. In: Kings Law Journals, 18 (2), 209-226.

Chadwick, R.; Berg, K. (2001): Solidarity and equity: new ethical frameworks for genetic databases. In: Nature Reviews Genetics, 2 (4), 318-321.

Chatziastros, A.; Drepper, J.; Semler, S. C. (2014): Big Data and Healthcare Analytics – datenschutzrechtliche Herausforderungen. In: mdi, 2/2014, 53-57.

Cisco (ed.) (2017): The Zettabyte Era: Trends and Analysis. https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.pdf [2017-10-17].

Condliffe, J. (2016): China turns big data into big brother. https://www.technologyreview.com/s/602987/china-turns-big-data-into-big-brother [2017-10-17].

Conley, A. et al. (2012): Sustaining privacy and open justice in the transition to online court records: A multidisciplinary inquiry. In: Maryland Law Review, 71 (3), 773-847.

Dabrock, P. (2012): Befähigungsgerechtigkeit. Ein Grundkonzept konkreter Ethik in fundamentaltheologischer Perspektive. Gütersloh.

Daniels, N. (1990): Equality of what: welfare, resources, or capabilities. In: Philosophy and Phenomenological Research, 50 (suppl.), 273-296.

Datta, A.; Tschantz, M. C. Datta, A.; (2015): Automated experiments on ad privacy settings. A tale of opacity, choice, and discrimination. In: Proceedings on Privacy Enhancing Technologies, 1/2015, 92-112.

Davis-Turak, J. et al. (2017): Genomics pipelines and data integration: challenges and opportunities in the research setting. In: Expert Review of Molecular Diagnostics, 17 (3), 225-237.

Determann, L.; Weigl, M. (2016): EU-US-Datenschutzschild und Alternativen für internationale Datentransfers. In: Europäische Zeitschrift für Wirtschaftsrecht, 27 (21), 811-816.

Deutscher Bundestag (ed.) (2017): Gesetzentwurf der Bundesregierung. Entwurf eines Gesetzes zur Anpassung des Datenschutzrechts an die Verordnung (EU) 2016/679 und zur Umsetzung der Richtlinie (EU) 2016/680 (Datenschutz-Anpassungs- und -Umsetzungsgesetz EU – DSAnpUG-EU). Bundestag printed paper 18/11325. http://dipbt.bundestag.de/doc/btd/18/113/1811325.pdf [2017-10-17].

Deutscher Bundestag (ed.) (2016): Antwort der Bundesregierung auf die Kleine Anfrage der Abgeordneten Maria Klein-Schmeink, Renate Künast, Dr. Konstantin von Notz, weiterer Abgeordneter und der Fraktion BÜNDNIS 90/DIE GRÜNEN. Drucksache 18/9058. Verhaltensbasierte Versicherungstarife – Apps und Wearables in der gesetzlichen Krankenversicherung. Bundestag printed paper 18/9243. http://dipbt.bundestag.de/doc/btd/18/092/1809243.pdf [2017-10-17].

Deutscher Bundestag (ed.) (2014): Kleine Anfrage der Abgeordneten Harald Weinberg, Kathrin Vogler, Jan Korte, weiterer Abgeordneter und der Fraktion DIE LINKE. Datensammlungen über Versicherte in der privaten Krankenversicherung. Bundestag printed paper 18/3633. http://dip21.bundestag.de/dip21/btd/18/036/1803633.pdf [2017-10-24].

Deutscher Bundestag (ed.) (2005): Gesetzentwurf der Fraktionen der CDU/CSU und SPD. Entwurf eines Gesetzes zur Verbesserung der Wirtschaftlichkeit in der Arzneimittelversorgung. Bundestag printed paper 16/194. http://dipbt.bundestag.de/doc/btd/16/001/1600194.pdf [2017-10-17].

Deutscher Ethikrat (ed.) (2016): Patientenwohl als ethischer Maßstab für das Krankenhaus. Berlin.

Deutscher Ethikrat (ed.) (2010): Human Biobanks for Research. Berlin.

Deutsches Institut für Vertrauen und Sicherheit im Internet (ed.) (2017): Digitalisierung – Deutsche fordern mehr Sicherheit. Hamburg.

Dijck, J. v. (2016): Big data, grand challenges: on digitization and humanities research. In: KWALON, 21 (1), 8-18.

Do, C. B. et al. (2011): Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. In: PLoS Genetics, 7 (6), art. no.: e1002141. DOI: 10.1371/journal.pgen.1002141.

Dorloff, A. (2017): Strafen und Belohnen per Big Data. http://www.deutschlandfunk.de/ueberwachung-in-china-strafen-und-belohnen-per-big-data.697.de.html?dram:article_id=387147 [2017-10-17].

Dorniok, D. (2015): Die Funktionalität eines Rechts auf Nichtwissen. Wiesbaden.

Drexl, J. (2017): Neue Regeln für die Europäische Datenwirtschaft? Ein Plädoyer für einen wettbewerbspolitischen Ansatz. In: Neue Zeitschrift für Kartellrecht, 5 (8), 339-344.

Duttge, G. (2010): Das Recht auf Nichtwissen in der Medizin. In: Datenschutz und Datensicherheit, 34 (1), 34.

Engeler, M.; Felber, W. (2017): Entwurf der ePrivacy-VO aus Perspektive der aufsichtsbehördlichen Praxis. Reguliert der Entwurf an der technischen Realität vorbei? In: Zeitschrift für Datenschutz, 7 (6), 251-257.

Ensthaler, J. (2016): Industrie 4.0 und die Berechtigung an Daten. In: Neure Juristische Wochenschrift, (69) 48, 3473-3478.

Ernst, S. (2017): Die Einwilligung nach der Datenschutzgrundverordnung. In: Zeitschrift für Datenschutz, 7 (3), 110-114.

Esteva, A. et al. (2017): Dermatologist-level classification of skin cancer with deep neural networks. In: Nature, 542 (7639), 115-118.

European Commission (ed.) (2017): Synopsis Report of the Public Consultation on the Evaluation and Review of the ePrivacy Directive. http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=40777 [2017-10-17].

European Commission (ed.) (2016): Draft Code of Conduct on Privacy for Mobile Health Applications. http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=16125 [2017-11-17]

European Commission (ed.) (2015): Manual on Borderline and Classification in the Community Regulatory Framework for Medical Devices. Version 1.17 (09-2015). https://www.emergogroup.com/sites/default/files/eu-manual-borderline-and-classification.pdf [2017-10-17].

European Commission (ed.) (2014): Grünbuch über Mobile-Health-Dienste („mHealth“). http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=5186 [2017-10-17].

European Commission (ed.) (2012a): High-Performance Computing: Europe's Place in a Global Race. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0045:FIN:EN:PDF [2017-11-14].

European Commission (ed.) (2012b): Guidelines on the Qualification and Classification of Stand Alone Software Used in Healthcare within the Regulatory Framework of Medical Devices. https://ec.europa.eu/docsroom/documents/17921/attachments/1/translations/en/renditions/native [2017-10-17].

European Commission (ed.) (2012c): Biobanks for Europe. A Challenge for Governance. http://www.irdirc.org/wp-content/uploads/2013/07/biobanks_for_Europe.pdf [2017-10-17].

European Data Protection Supervisor (ed.) (2015): Mobile Health. Reconciling technological innovation with data protection https://edps.europa.eu/sites/default/files/publication/15-05-21_mhealth_en_0.pdf [2021-08-19].

Fachforum “Digitalisierung und Gesundheit” im Hightech-Forum (ed.) (2017): Die Zukunft der Medizin ist digital. Szenario zur Digitalisierung in der Gesundheitsversorgung. Munich.

Fähnrich, C. et al. (2015): Surveillance and outbreak response management system (SORMAS) to support the control of the ebola virus disease outbreak in West Africa. In: Eurosurveillance, 20 (12), art. no.: pii=21071. DOI: 10.2807/1560-7917.ES2015.20.12.21071.

Feinberg, J. (1989): The Moral Limits of the Criminal Law. Volume 3: Harm to Self. Oxford; NewYork; Toronto.

Fetzer, T. (2015): Plädoyer für ein neues Datenrecht. In: MultiMedia und Recht, 16 (12), 777-778.

Fezer, K.-H. (2017a): Dateneigentum der Bürger. Ein originäres Immaterialgüterrecht sui generis an verhaltensgenerierung Informationsdaten der Bürger. In: Zeitschrift für Datenschutz, 7 (3), 99-104.

Fezer, K.-H. (2017b): Dateneigentum. Theorie des immaterialgüterrechtlichen Eigentums an verhaltensgenerierten Personendaten der Nutzer als Datenproduzenten. In: MultiMedia und Recht, 18 (1), 3-5.

Floridi, L. (2014): The Fourth Revolution – How the Infosphere is Reshaping Human Reality. Oxford.

Forst, R. (2007): Das Recht auf Rechtfertigung: Elemente einer konstruktivistischen Theorie der Gerechtigkeit. Berlin.

Forst, R. (1996): Kontexte der Gerechtigkeit. Politische Philosophie jenseits von Liberalismus und Kommunitarismus. Frankfurt am Main.

Foucault, M. (1976): Überwachen und Strafen. Die Geburt des Gefängnisses. Frankfurt am Main.

Frackowiak, R.; Ailamaki, A.; Kherif, F. (2016): Federating and Integrating What We Know About the Brain at All Scales: Computer Science Meets the Clinical Neurosciences. Cham.

Francke, R.; Hart, D. (2008): Bewertungskriterien und -methoden nach dem SGB V. In: MedizinRecht, 26 (1), 2-24.

Friedrichsen, M.; Bisa, P.-J. (ed.) (2016): Digitale Souveränität. Vertrauen in der Netzwerkgesellschaft. Wiesbaden.

Gärtner, A. (2010): MDD 2007/47/EG: Software als Medizinprodukt. http://www.e-health-com.eu/fileadmin/user_upload/dateien/Downloads/Gaertner-Medizinprodukte-Gesetz.pdf [2017-10-17].

Gassner, U. (2016): Software als Medizinprodukt – zwischen Regulierung und Selbstregulierung. In: Medizin Produkte Recht, 16 (4), 109-115.

Gassner, U. (2015): MedTech meets M-Health. In: Medizin Produkte Recht, 15 (3), 73-82.

Gellman, R.; Dixon, P. (2016): Failures of privacy self-regulation in the United States. In: D. Wright; P. De Hert (ed.): Enforcing Privacy. Regulatory, Legal and Technological Approaches. Cham, 53-77.

Gerhardt, V. (1999): Selbstbestimmung. Das Prinzip der Individualität. Stuttgart.

Gesang, B. (2013): Rechte und Autonomie im Utilitarismus. In: H. Greif; M. G. Weiss (ed.): Ethics, Society, Politics: Proceedings of the 35th International Wittgenstein Symposium, Kirchberg on the Wechsel, Austria 2012 (Publications of the Austrian Ludwig Wittgenstein Society – New Series, volume 20). Berlin, 227-240.

Gesang, B. (2003): Eine Verteidigung des Utilitarismus. Stuttgart.

Gesang, B. (2000): Der Nutzenbegriff des Utilitarismus. In: Erkenntnis, 52 (3), 373-401.

Gethmann, C. F. (1996): Heilen: Können und Wissen. Zu den philosophischen Grundlagen der wissenschaftlichen Medizin. In: J. P. Beckmann (ed.): Fragen und Probleme einer medizinischen Ethik. Berlin, 68-93.

GKV-Spitzenverband (ed.) (2014): Stellungnahme des GKV-Spitzenverbandes vom 27.06.2014 zur öffentlichen Konsultation der Europäischen Kommission zum Grünbuch über Mobile-Health-Dienste (SWD (2014) 135 final). http://dsv-europa.de/lib/02_Positionspapiere/2014_GKV-SV_mHealth_de.pdf [2017-10-17].

Gossen, H.; Schramm, M. (2017): Das Verarbeitungsverzeichnis der DS-GVO. Ein effektives Instrument zur Umsetzung der neuen unionsrechtlichen Vorgaben. In: Zeitschrift für Datenschutz, 7 (1), 7-13.

Gräßer, F. et al. (2017): Therapy decision support based on recommender system methods. In: Journal of Healthcare Engineering, art. no.: 8659460. DOI: 10.1155/2017/8659460.

Greenstein, S.; Nagle, F. (2014): Digital dark matter and the economic contribution of Apache. In: Research Policy, 43 (4), 623-631.

Grimm, O.; Maiß, S. (2015): Bewerbersuche, Datenschutz und Headhunting. Social-Media-Recruiting. In: Arbeit und Arbeitsrecht, 70 (5), 270-272.

Habl, C. et al. (2016): Study on Big Data in Public Health, Telemedicine and Healthcare. Final Report. https://ec.europa.eu/health/sites/health/files/ehealth/docs/bigdata_report_en.pdf. [2017-10-17].

Halpern, D. (2015): Inside the Nudge Unit. How Small Changes Can Make a Big Difference. London.

Hanfeld, M. (2017): Profitieren am Ende wieder nur Google und Co.? http://www.faz.net/aktuell/wirtschaft/die-reform-der-eprivacy-nutzt-am-ende-nur-google-und-co-15266269.html [2017-10-17].

Hart, D. (2000): Evidenz-basierte Medizin und Gesundheitsrecht – Überlegungen zu rechtlichen Konsequenzen der Verwissenschaftlichung der Medizin. In: MedizinRecht, 18 (1), 1-5.

Hauner, W. (2016): Big Data Analysis @ Munich Re. https://www.munichre.com/site/daktylos/get/documents_E912162360/mr/assets.dakylos/Documents/presentation/life-forum/Big-Data-Analytics-Munich-Re.pdf [2017-10-17].

Heidbrink, L.; Langbehn, C.; Loh, J. (2017): Handbuch Verantwortung. Wiesbaden.

Heimhalt, D.; Rehmann, W. (2014): Gesundheits- und Patienteninformationen via Apps. In: Medizin Produkte Recht, 8 (6), 197-205.

Helfrich, M. (2017): Art. 22. Automatisierte Entscheidungen im Einzelfall einschließlich Profiling. In: G. Sydow (ed.): Europäische Datenschutzgrundverordnung. Baden-Baden, 559-574.

Heller, C. (2011): Post-Privacy. Prima leben ohne Privatsphäre. Munich.

Hellgardt, A. (2016): Regulierung und Privatrecht. Tübingen.

Henn, W. (2016): Prädiktive und individualisierte Genmedizin: Ethische und unternehmenspolitische Herausforderungen. Zeitschrift für Versicherungswesen, 8 (18), 563-566.

Hermstrüwer, Y. (2016): Informationelle Selbstgefährdung. Tübingen.

Heuvel, M. v. d. (2016): Gefährlicher Schatz. https://www.deutsche-apotheker-zeitung.de/news/artikel/2016/03/30/gefahrlicher-schatz [2017-11-03].

Hey, T.; Tansley, S.; Tolle, K. (Hg.) (2009): The Fourth Paradigm. Data Intensive Scientific Discovery. Redmond.

Hobbes, T. (1999): Leviathan oder Stoff, Form und Gewalt eines kirchlichen und bürgerlichen Staates. Herausgegeben von I. Fetscher. Frankfurt am Main.

Hoffmann-Riem, W. (2000): Verwaltungsrecht in der Informationsgesellschaft – Einleitende Problemskizze. In: W. Hoffmann-Riem; E. Schmidt-Aßmann (ed.): Verwaltungsrecht in der Informationsgesellschaft. Baden-Baden, 9-58.

Hofmann, J. M.; Johannes, P. C. (2017): DS-GVO: Anleitung zur autonomen Auslegung des Personenbezugs. In: Zeitschrift für Datenschutz, 7 (5), 221-226.

Horak, P. et al. (2017): Precision oncology based on omics data: the NCT Heidelberg experience. In: International Journal of Cancer, 141 (5), 877-886.

Hyman, J.; Steward, H. (ed.) (2004): Agency and Action. Cambridge.

Institute for Government (ed.) (2010): Mindspace. Influencing Behavior through Public Policy. https://www.instituteforgovernment.org.uk/sites/default/files/publications/MINDSPACE.pdf [2017-10-17].

International Bioethics Committee (2017): Report of the IBC on Big Data and Health. http://unesdoc.unesco.org/images/0024/002487/248724E.pdf [2017-11-17].

Jones, D. T. W. et al. (2013): Recurrent somatic alterations of FGFR1 and NTRK2 in pilocytic astrocytoma. In: Nature Genetics, 45 (8), 927-932.

Kamnitsas, K. et al. (2017): Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. https://arxiv.org/pdf/1612.08894.pdf [2017-10-17].

Kampert, D. (2017): Art. 8. Bedingungen für die Einwilligung eines Kindes in Bezug auf Dienste der Informationsgesellschaft. In: G. Sydow (ed.): Europäische Datenschutzgrundverordnung. Baden-Baden, 381-388.

Kane, R. (ed.) (2011): The Oxford Handbook of Free Will (2nd ed.). Oxford; New York.

Kant, I. (1903): Grundlegung zur Metaphysik der Sitten. In: Kant, I.: Gesammelte Schriften. Volume 4 (I/4): Kritik der reinen Vernunft. Prolegomena. Grundlegung zur Metaphysik der Sitten. Metaphysische Anfangsgründe der Naturwissenschaft. Berlin, 385-463.

Karaalp, R. N. (2016): Der Schutz von Patientendaten für die medizinische Forschung in Krankenhäusern. Wiesbaden.

Katko, P.; Babaei-Beigi, A. (2014): Accountability statt Einwilligung? – Führt Big Data zum Paradigmenwechsel im Datenschutz? In: MultiMedia und Recht, 15 (6), 360-364.

Kaye, J. et al. (2015): Dynamic consent: a patient interface for twenty-first century research networks. In: European Journal of Human Genetics, 23 (2), 141-146.

Keppeler, L.; Berning, W. (2017): Technische und rechtliche Probleme bei der Umsetzung der DS-GVO-Löschpflichten. Anforderungen an Löschkonzepte und Datenbankstrukturen. In: Zeitschrift für Datenschutz, 7 (7), 314-318.

Keßler, O. (2017): Intelligente Roboter – neue Technologien im Einsatz. In: MultiMedia und Recht, 18 (9), 589-594.

Kettemann, M. C. (2015): Völkerrecht in den Zeiten des Netzes. Perspektiven auf den Schutz von Grund- und Menschenrechten in der Informationsgesellschaft zwischen Völkerrecht, Europarecht und Staatsrecht. Bonn.

Kim, Y.; Huang, J.; Emery, S. (2016): Garbage in, garbage out: data collection, quality assessment and reporting standards for social media data use in health research, infodemiology and digital disease detection. In: Journal of Medical Internet Research, 18 (2), art. no.: 41. DOI: 10.2196/jmir.4738.

Kitchin, R. (2014): Big data, new epistemologies and paradigm shifts. In: Big Data & Society, 1 (1), 1-12.

Kitsiou, S. et al. (2017): Effectiveness of mHealth interventions for patients with diabetes: An overview of systematic reviews. In: PloS ONE, 12 (3), art. no.: e0173160. DOI: 10.1371/journal.pone.0173160.

Klein, R. A. (2016): Depotenzierung der Souveränität. Tübingen.

Klement, J. H. (2017): Öffentliches Interesse an Privatheit. Das europäische Datenschutzrecht zwischen Binnenmarkt, Freiheit und Gemeinwohl. In: JuristenZeitung, 72 (4), 161-170.

Kneer, G. (2012): Die Analytik der Macht bei Michel Foucault. In: P. Imbusch (ed.): Macht und Herrschaft. Wiesbaden, 265-283.

Kollek, R. (2012): Systembiologie als Nexus zwischen Genen und Gesundheit? TA-Implikationen konzeptioneller Innovation in den Lebenswissenschaften. In: Technikfolgenabschätzung – Theorie und Praxis, 21 (2), 21-28.

Koppernock, M. (1997): Das Grundrecht auf bioethische Selbstbestimmung. Zur Rekonstruktion des allgemeinen Persönlichkeitsrechts. Baden-Baden.

Krajewski, J. et al. (2014): A phonetic approach for detecting sleepiness from speech in simulated air-traffic controller-communication. In: D. d. Waard et al. (ed.): Human Factors of Systems and Technology. Maastricht, 147-155.

Kramer, A. D. I.; Guillory, J. E.; Hancock, J. T. (2014): Experimental evidence of massive-scale emotional contagion through social networks. In: Proceedings of the National Academy of Sciences of the United States of America, 111 (24), 8788-8790.

Kreße, B. (2017): Artikel 82. Haftung und Recht auf Schadensersatz. In: G. Sydow (ed.): Europäische Datenschutzgrundverordnung. Baden-Baden, 1260-1268.

Krohm, N. (2016): Abschied vom Schriftformgebot der Einwilligung – Lösungsvorschläge und künftige Anforderungen. In: Zeitschrift für Datenschutz, 6 (8), 368-373.

Krügel, T. (2017): Das personenbezogene Datum nach der DS-GVO. Mehr Klarheit und Rechtssicherheit? In: Zeitschrift für Datenschutz, 7 (10), 455-459.

Krüger, P.-L. (2016): Datensouveränität und Digitalisierung. In: Zeitschrift für Rechtspolitik, 49 (7), 190-192.

Kucklick, C. (2016): Die granulare Gesellschaft. Wie das Digitale unsere Wirklichkeit auflöst. Berlin.

Kühling, J.; Seidel, C. (2015): Grundlagen – Allgemeiner Teil. In: T. Kingreen; J. Kühling (ed.): Gesundheitsdatenschutzrecht. Baden-Baden, 29-185.

Ladeur, K.-H. (2016): „Big Data" im Gesundheitsrecht – Ende der „Datensparsamkeit"? In: Datenschutz und Datensicherheit, 40 (6), 360-364.

Laney, D. (2001): 3-D Data Management: Controlling Data Volume, Velocity and Variety. https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf [2017-10-17].

Langanke, M. et al. (2013): Gesundheitliche Eigenverantwortung im Kontext Individualisierter Medizin. In: Ethik in der Medizin, 25 (3), 243-250.

Langkafel, P. (2015): Auf dem Weg zum Dr. Algorithmus? Potenziale von Big Data in der Medizin. In: Aus Politik und Zeitgeschichte, 65 (11-12), 27-32.

Langkafel, P. (ed.) (2014): Big Data in Medizin und Gesundheitswirtschaft. Heidelberg.

Lauss, G. et al. (2011): Embracing complexity and uncertainty: An analysis of three orders of ELSA research on biobanks. In: Genomics, Society and Policy, 7 (1), 47-64.

LeCun, Y.; Bengio, Y.; Hinton, G. (2015): Deep learning. In: Nature, 521 (7553), 436-444.

Lek, M. et al. (2016): Analysis of protein-coding genetic variation in 60,706 humans. In: Nature, 536 (7616), 285-291.

Leong, L. et al. (2017): Magic Quadrant for Cloud Infrastructure as a Service, Worldwide. Ed. by Gartner. https://www.gartner.com/doc/reprints?id=1-2G2O5FC&ct=150519&st=sb [2017-10-17].

Lewinski, K. v. (2016): Privacy Shield – Notdeich nach dem Pearl Harbor für die transatlantischen Datentransfers. In: Europarecht, 13 (4), 405-421.

Lewinski, K. v.; Herrmann, C. (2016): Cloud vs. Cloud – Datenschutz im Binnenmarkt. Verantwortlichkeit und Zuständigkeit bei grenzüberschreitender Datenverarbeitung. In: Zeitschrift für Datenschutz, 6 (10), 467-474.

Litjens, G. (2017): A survey on deep learning in medical image analysis. In: Medical Image Analysis, 42, 60-88.

Livingston, K. M. et al. (2013): Representing annotation compositionality and provenance for the semantic web. In: Journal of Biomedical Semantics, 4, art. no.: 38. DOI: 10.1186/2041-1480-4-38.

Lob-Hüdepohl, A. (2007): Schwierige Willensbekundung. Garantieren Patientenverfügungen würdevolles Sterben? In: Herder Korrespondenz, 61 (2), 83-87.

Lovett, F. (2007): Power. In: R. E. Goodin; P. Pettit; T. W. Pogge (ed.): A Companion to Contemporary Political Philosophy (2nd ed.). Malden (MA), 709-718.

Lucht, M.; Boeker, M.; Kramer, U. (2017): Gesundheits- und Versorgungsapps. Ed. by Universitätsklinikum Freiburg. http://www.tk.de/centaurus/servlet/contentblob/724464/Datei/143235/Studie-Gesundheits-und-Versorgungs-Apps.pdf [2017-10-17].

Lukes, S. (2005): Power. A Radical View (2nd ed.). Basingstoke; New York.

Martini, M. (2016): Do it yourself im Datenschutzrecht. In: Neue Zeitschrift für Verwaltungsrecht, 35 (6), 353-354.

Martini, M. (2014): Big Data als Herausforderung für den Persönlichkeitsschutz und das Datenschutzrecht. In: Deutsches Verwaltungsblatt, 129 (23), 1481-1489.

Martini, M.; Nink, F. (2017): Wenn Maschinen entscheiden … – vollautomatisierte Verwaltungsverfahren und der Persönlichkeitsschutz. In: Neue Zeitschrift für Verwaltung, 36 (10), 1-14.

Mayer-Schönberger, V.; Cukier, K. (2013): Big Data. A Revolution That Will transform How We Live, Work, and Think. Boston; London.

McGoogan, C. (2017): NHS illegally handed Google firm 1.6m patient records, UK data watchdog finds. http://www.telegraph.co.uk/technology/2017/07/03/googles-deepmind-nhs-misused-patient-data-trial-watchdog-says [2017-11-17].

Merkel, R. et al. (2007): Intervening in the Brain. Changing Psyche and Society. Berlin; Heidelberg.

Mill, J. S. (1859): On Liberty. London.

Molnár-Gábor, F.; Kaffenberger, L. (2017): EU-US-Privacy-Shield – ein Schutzschild mit Löchern? In: Zeitschrift für Datenschutz, 7 (1), 18-24.

Montjoye, Y.-A. de et al. (2013): Unique in the crowd. The privacy bounds of human mobility. In: Scientific Reports, 3, art. no.: 1376. DOI: 10.1038/srep01376.

Mooy, M. D. (2017): Rethinking Privacy Self-Management and Data Sovereignty in the Age of Big Data. Ed. by Bertelsmann Stiftung. Gütersloh.

Mostert, M. et al. (2016): Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach. In: European Journal of Human Genetics, 24 (7), 956-960.

Mota, N. B. et al. (2012): Speech graphs provide a quantitative measure of thought disorder in psychosis. In: PLoS One, 7 (4), art. no.: e34928. DOI: 10.1371/journal.pone.0034928.

Müller, H.; Hanbury, A. (2016): Forschungsanwendungen in der digitalen Radiologie. In: Der Radiologe, 56 (2), 176-180.

Nakamoto, S. (2009): Bitcoin. A Peer-to-Peer Electronic Cash System. https://bitcoin.org/bitcoin.pdf [2017-10-17].

Nationaler Ethikrat (ed.) (2007): Prädiktive Gesundheitsinformationen beim Abschluss von Versicherungen. Berlin.

Nationaler Ethikrat (ed.) (2005): Prädiktive Gesundheitsinformationen bei Einstellungsuntersuchungen. Berlin.

Nationaler Ethikrat (ed.) (2004): Biobanken für die Forschung. Berlin.

Nationales Aktionsbündnis für Menschen mit Seltenen Erkrankungen (ed.) (2016): Zwischenbericht zur Umsetzung des Nationalen Aktionsplans für Menschen mit Seltenen Erkrankungen. http://www.achse-online.de/de/was_tut_ACHSE/namse/pdf/nationaler_Aktionsplan/namse_monitoringbericht_2016.pdf [2017-10-17].

Nicholson Price II, W. (2015): Black-box medicine. In: Havard Journal of Law & Technology, 28 (2), 419-468.

Nissenbaum, H. (2011): A contextual approach to privacy online. In: Daedalus, 140 (4), 32-48.

Nissenbaum, H. (2009): Privacy in Context: Technology, Policy, and the Integrity of Social Life. Stanford.

Nussbaum, M. (2015): Fähigkeiten schaffen: neue Wege zur Verbesserung menschlicher Lebensqualität. Freiburg im Breisgau.

Nussbaum, M. (2014): Die Grenzen der Gerechtigkeit. Behinderung, Nationalität und Spezienszugehörigkeit. Berlin.

Nussbaum, M. (1998): Gerechtigkeit oder Das gute Leben. Frankfurt am Main.

Obermeyer, Z.; Emanuel, E. J. (2016): Predicting the future – big data, machine learning, and clinical medicine. In: New England Journal of Medicine, 375 (13), 1216-1219.

OECD (ed.) (2013): The OECD Privacy Framework. http://oecd.org/sti/ieconomy/oecd_privacy_framework.pdf [2017-10-17].

Oen, R. (2009): Software als Medizinprodukt. In: Medizin Produkte Recht, 3 (2), 55-57.

Ohm, P. (2015): Sensitive information. In: Southern California Law Review, 88 (5), 1125-1196.

Ohm, P. (2010): Broken promises of privacy: responding to the surprising failure of anonymization. In: UCLA Law Review, 57 (6), 1701-1777.

Ortner, R.; Daubenbüchel, F. (2016): Medizinprodukte 4.0 – Haftung, Datenschutz, IT-Sicherheit. In: Neue Juristische Wochenschrift, 69 (40), 2918-2924.

Paal, B.; Hennemann, M. (2017): Big Data im Recht – Wettbewerbs- und daten(schutz)rechtliche Herausforderungen. In: Neue Juristische Wochenschrift, 70 (24), 1697-1701.

Pannenbecker, A. (2013): § 14 Grundzüge des Arzneimittel- und Medizinprodukterechts. In: M. Terbille; T. Clausen; J. Schroeder-Printzen (ed.): Münchener Anwaltshandbuch Medizinrecht (2nd ed.). Munich, 1165-1283.

Pearl, J. (2010): An introduction to causal inference. In: The International Journal of Biostatistics, 6 (2), 1-59.

Pelz, W. (2017): SWOT-Analyse. Definition, Beispiele und Vorlagen zum Erstellen einer SWOT-Analyse. http://www.wpelz.de/swot-analyse/SWOT-Analyse.pdf [2017-10-17].

Pillay, N. (2014): The right to privacy in the digital age. Report of the Office of the United Nations High Commissioner for Human Rights. UN document A/HRC/27/37. http://www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session27/Documents/A-HRC-27-37_en.doc [2017-10-17].

Ploug, T.; Holm, S. (2016): Meta consent – a flexible solution to the problem of secondary use of health data. In: Bioethics, 30 (9), 721-732.

Poelzig, D. (2012): Normdurchsetzung durch Privatrecht. Tübingen.

Polonetsky, J.; Tene, O. (2013): Privacy and big data. Making ends meet. In: Stanford Law Review, 66, art. no.: 25. SSRN: 2628412.

Prainsack, B.; Buyx, A. (2017): Solidarity in Biomedicine and Beyond. Cambridge.

Prainsack, B.; Buyx, A. (2016): Das Solidaritätsprinzip. Ein Plädoyer für eine Renaissance in Medizin und Bioethik. Frankfurt am Main.

Radic, D. et al. (2016): Big Data im Krankenversicherungsmarkt. Relevanz, Anwendungen, Chancen und Hindernisse. http://s.fhg.de/krankenversicherung [2017-11-17].

Rat für Informationsinfrastrukturen (ed.) (2017): Schritt für Schritt – oder: Was bringt wer mit? Ein Diskussionsimpuls zu Zielstellung und Voraussetzungen für den Einstieg in die Nationale Forschungsdateninfrastruktur (NFDI). http://www.rfii.de/?wpdmdl=2269 [2017-10-17].

Rawls, J. (1972): A Theory of Justice. Oxford.

Regalado, A. (2017): EmTech: Illumina Says 228,000 Human Genomes Will Be Sequenced This Year. https://www.technologyreview.com/s/531091/emtech-illumina-says-228000-human-genomes-will-be-sequenced-this-year [2017-10-17].

Reich, N. (2014): Fehlerhaftigkeit von Medizinprodukten. Europäische Zeitschrift für Wirtschaftsrecht, 25 (23), 898-900.

Reinsel, D.; Gantz, J.; Rydning, J. (2017): Data Age 2025. The Evolution of Data to Life-Critical. Ed. by IDC. http://www.seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf [2017-10-17].

Rhees R. (ed.) (1969): Schriften von Ludwig Wittgenstein. Band I: Tractatus logico-philosophicus; Tagebücher; Philosophische Untersuchungen. Frankfurt am Main.

Richter, G.; Buyx, A. (2016): Breite Einwilligung (broad consent) zur Biobank-Forschung – die ethische Debatte. In: Ethik in der Medizin, 28 (4), 311-325.

Riechert, A. (2016): Stellungnahme zu rechtlichen Aspekten eines Einwilligungsassistenten. https://stiftungdatenschutz.org/fileadmin/Redaktion/Bilder/Abschluss_Studie_30032017/stiftungdatenschutz_Stellungnahme_Rechtliche_Aspekte_eines_Einwilligungsassistenten_Anhang_1_final.pdf [2017-10-17].

Roberto, C. A.; Kawachi, I. (2015): Behavioral Economics and Public Health. Oxford.

Rorty, R. (1989): Contingency, Irony and Solidarity. Cambridge.

Roessler, J.; Eilan, N. (ed.) (2003): Agency and Self-Awareness. Issues in Philosophie and Psychology. Oxford.

Rössler, B. (2001): Der Wert des Privaten. Berlin.

Roßnagel, A.; Nebel, M. (2015): (Verlorene) Selbstbestimmung im Datenmeer – Privatheit im Zeitalter von Big Data. Datenschutz und Datensicherheit, 39 (7), 455-459.

Rübsamen, K. (2015): Rechtliche Rahmenbedingungen für mobileHealth. In: MedizinRecht, 33 (7), 485-491.

Rudolf, W. (2011): § 90 - Recht auf informationelle Selbstbestimmung. In: D. Merten; H.-J. Papier (ed.): Handbuch der Grundrechte. Volume 4. Heidelberg, 233-290.

Samuelson, P. (1975): Foundations of Economic Analysis (10th ed.). Cambridge.

Schaar, K. (2017): Anpassung von Einwilligungserklärungen für wissenschaftliche Forschungsprojekte. Die informierte Einwilligung nach der DS-GVO und den Ethikrichtlinien. In: Zeitschrift für Datenschutz, 7 (5), 213-220.

Schaar, P. (2014): Überwachung total. Wie wir in Zukunft unsere Daten schützen. Berlin.

Schantz, P. (2016): Die Datenschutz-Grundverordnung – Beginn einer neuen Zeitrechnung im Datenschutzrecht. In: Neue Juristische Wochenschrift, 69 (26), 1841-1847.

Scherer, S. et al. (2013): Investigating voice quality as a speaker-independent indicator of depression and PTSD. https://pdfs.semanticscholar.org/d7dd/8fc189f93d9a1471458290 4dd9542f952f32.pdf [2017-10-17].

Schmitz, B.; Dall'Armi, J. v. (2017): Datenschutz-Folgenabschätzung – verstehen und anwenden. Wichtiges Instrument zur Umsetzung von Privacy by Design. In: Zeitschrift für Datenschutz, 7 (2), 57-63.

Schneider, J. (2017): Schließt Art. 9 DS-GVO die Zulässigkeit der Verarbeitung bei Big Data aus? Überlegungen, wie weit die Untersagung bei besonderen Datenkategorien reicht. In: Zeitschrift für Datenschutz, 7 (7), 303-307.

Schröder, P. (2011): Politische Strategien (2nd ed.). Herausgegeben von der Friedrich-Naumann-Stiftung für die Freiheit. Potsdam.

Schroeder, A. (2015): Das Recht auf Nichtwissen im Kontext prädiktiver Gendiagnostik. Eine Studie zum ethisch verantworteten Umgang mit den Grenzen des Wissens. Wiesbaden.

Schwartz, P. M.; Solove, D. J. (2011): The PII problem: privacy and a new concept of personally identifiable information. In: New York University Law Review, 86 (6), 1814-1894.

Seemann, M. (2011): Vom Kontrollverlust zur Filtersouveränität. In: Heinrich-Böll-Stiftung (ed.): #public_life. Digitale Intimität, die Privatsphäre und das Netz. Berlin, 74-80.

Seife, C. (27 November 2013): 23andMe is terrifying, but not for the reasons the FDA thinks. https://www.scientificamerican.com/article/23andme-is-terrifying-but-not-for-the-reasons-the-fda-thinks [2017-10-17].

Sen, A. (2009): The Idea of Justice. Cambridge.

Sen, A. (1979): Equality of what? In: S. McMurrin (ed.): The Tanner Lectures on Human Values. Cambridge, 195-220.

Simitis, S. (2014a): § 4a. Einwilligung. In: S. Simitis (ed.): Bundesdatenschutzgesetz (8th ed.). Baden-Baden, 470-503.

Simitis, S. (2014b): § 28. Datenerhebung und -speicherung für eigene Geschäftszwecke. In: S. Simitis (ed.): Bundesdatenschutzgesetz (8th ed.). Baden-Baden, 1182-1271.

Sivridis, A.; Seidel, C.; Kühling, J. (2015): Datenschutzrecht (3rd ed.). Heidelberg.

Smart, J. J. C. (1961): An Outline of a System of Utilitarian Ethics. Melbourne.

Smart, J. J. C.; Williams, B. (1973): Utilitarianism For and Against. Cambridge.

Solove, D. J. (2011): Nothing to Hide. The False Tradeoff Between Privacy and Security. New Haven.

Solove, D. J. (2008): Understanding Privacy. Cambridge.

Solove, D. J. (2007): "I've got nothing to hide" and other misunderstandings of privacy. In: San Diego Law Review, 44 (4), 745-772.

Solove, D. J.; Hartzog, W. (2014): The FTC and the new common law of privacy. In: Columbia Law Review, 114 (3), 583-676.

Specht, L. (2017): Daten als Gegenleistung – Verlangt die Digitalisierung nach einem neuen Vertragstypus? In: JuristenZeitung, 72 (15-16), 763-770.

Spielkamp, M. (2017): Sind Algorithmen die besseren Richter? In: Technology Review, 8/2017, 36-37.

Stallberg, C. (2010): Evidenzbasierte Medizin als Rechtsbegriff – Funktion, Inhalte, Grenzen. In: Pharma Recht, 32 (1), 5-12.

Stegmüller, W. (1983): Probleme und Resultate der Analytischen Philosophie. Volume I (2nd ed.). Berlin.

Steinsbekk, K. S.; Kåre Myskja, B.; Solberg, B. (2013): Broad consent versus dynamic consent in biobank research: Is passive participation an ethical problem? In: European Journal of Human Genetics, 21 (9), 897-902.

Stephens, Z. D. et al. (2015): Big Data. Astronomical or genomical? In: PLoS Biology, 13 (7), art. no.: e1002195. DOI: 10.1371/journal.pbio.1002195.

Stone, M. L. (2014): Big Data for Media. Ed. by Reuters Institute for the Study of Journalism and University of Oxford. http://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-04/Big%20Data%20For%20Media_0.pdf [2017-10-17].

Strategy&; PricewaterhouseCoopers (ed.) (2016): Weiterentwicklung der eHealth-Strategie. Sudie im Auftrag des Bundesministeriums für Gesundheit. http://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/E/eHealth/BMG-Weiterentwicklung_der_eHealth-Strategie-Abschlussfassung.pdf [2017-10-17].

Struppek, D. (2010): Patientensouveränität im Pflegeheim – Möglichkeiten und Grenzen aus der Sicht von hochaltrigen, mehrfach erkrankten Pflegeheimbewohnern, ihren Ärzten, Pflegekräften und privaten Bezugspersonen. Berlin.

Taigman, Y. (2014): Web-Scale Training for Face Identification. http://www.cs.tau.ac.il/~wolf/deeplearningmeeting/pdfs/deepface_masterclass.pdf [2017-10-17].

Taigman, Y. et al. (2014): Deepface: Closing the Gap to Human-Level Performance in Face Verification. https://www.cs.toronto.edu/~ranzato/.../taigman_cvpr14.pdf [2017-10-17].

Tamm, M. (2011): Verbraucherschutzrecht. Europäisierung und Materialisierung des deutschen Zivilrechts und die Herausbildung eines Verbraucherschutzprinzips. Tübingen.

Tene, O.; Polonetsky, J. (2012): Privacy in the age of big data: a time for big decisions. In: Stanford Law Review Online, 64 (25), 63-69.

Teramoto, A. et al. (2017): Automated classification of lung cancer types from cytological images using deep convolutional neural networks. In: BioMed Research International, art no.: 4067832. DOI: 10.1155/2017/4067832.

Thaler, R. H.; Sunstein, C. R. (2008): Nudge. Wie man kluge Entscheidungen anstößt. Berlin.

Thorbom, N. (2015): Datenschutz in der medizinischen Forschung. In: T. Kingreen; J. Kühling (ed.): Gesundheitsdatenschutzrecht. Baden-Baden, 344-371.

Thüsing, G.; Schmidt, M.; Forst, G. (2017): Das Schriftformerfordernis der Einwilligung nach § 4a BDSG im Pendelblick zu Art. 7 DS-GVO. In: Recht der Datenverarbeitung, 33 (3), 116-222.

Thouvenin, F. (2017): Wem gehören meine Daten? Zu Sinn und Nutzen einer Erweiterung des Eigentumsbegriffs. In: Schweizerische Juristen-Zeitung, 113 (2), 21-32.

Towfigh, E. V.; Ulrich, J. (2017): Artikel 44. Allgemeine Grundsätze der Datenübermittlung. In: G. Sydow (ed.): Europäische Datenschutzgrundverordnung. Baden-Baden. 862-872.

Tsanas, A. et al. (2012): Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. In: IEEE Transactions on Biomedical Engineering, 59 (5), 1264-1271.

Ulbricht, M.-R.; Weber, K. (2017): Adieu Einwilligung? Neue Herausforderungen für die informationelle Selbstbestimmung im Angesicht von Big Data-Technologien. In: M. Friedewald; J. Lamla; A. Roßnagel (ed.): Informationelle Selbstbestimmung im digitalen Wandel. Wiesbaden, 265-286.

Verbraucherzentrale Nordrhein-Westfalen (ed.) (2015): Bonusprogramme der gesetzlichen Krankenkassen. https://www.verbraucherzentrale.nrw/sites/default/files/migration_files/media236794A.pdf [2017-10-24].

Vodafone Institute for Society and Communications (ed.) (2016): Big Data. A European Survey on the Opportunities and Risks of Data Analytics. http://www.vodafone-institut.de/wp-content/uploads/2016/01/VodafoneInstitute-Survey-BigData-en.pdf [2017-10-17].

Vossenkuhl, C. (2013): Der Schutz genetischer Daten. Unter besonderer Berücksichtigung des Gendiagnostikgesetzes. Berlin.

Walzer, M. (1983): Spheres of Justice: A Defense of Pluralism and Equality. New York.

Wartenberg, T. E. (1990): The Forms of Power: From Domination to Transformation. Philadelphia.

Warren, S. D.; Brandeis, L. D. (1890): The right to privacy. In: Harvard Law Review, 4 (5), 193-220.

Weber, M. (1978): Economy and Society. An Outline of Interpretive Sociology. Berkeley.

Wee, R. (2013): Dynamic consent in the digital age of biology. In: Journal of Primary Health Care, 5 (3), 259-261.

Weichert, T. (2014a): Medizinisches Big Data und Datenschutz. In: P. Langkafel (ed.): Big Data in Medizin und Gesundheitswirtschaft. Heidelberg, 161-174.

Weichert, T. (2014b): Big Data, Gesundheit und der Datenschutz. In: Datenschutz und Datensicherheit, 38 (12), 831-838.

Weichert, T. (2013): Big Data und Datenschutz. Chancen und Risiken einer neuen Form der Datenanalyse. In: Zeitschrift für Datenschutz, 3 (6), 251-259.

Weilert, A. K. (2015): Gesundheitsverantwortung zwischen Markt und Staat. Baden-Baden.

Weitzel, T. et al. (2016): Active Sourcing und Social Recruiting. https://www.uni-bamberg.de/fileadmin/uni/fakultaeten/wiai_lehrstuehle/isdl/Recruiting_Trends_2016_-_Active_Sourcing_und_Social_Recruiting_v_WEB.PDF [2017-10-17].

Werkmeister, C.; Brandt, E. (2016): Datenschutzrechtliche Herausforderungen für Big Data. In: Computer und Recht, 32 (4), 233-238.

Wiebe, A. (2017): Schutz von Maschinendaten durch das sui-generis-Schutzrecht für Datenbanken. In: Gewerblicher Rechtsschutz und Urheberrecht, 119 (4), 338-345.

Wiebe, A.; Schur, N. (2017): Ein Recht an industriellen Daten im verfassungsrechtlichen Spannungsverhältnis zwischen Eigentumsschutz, Wettbewerbs- und Informationsfreiheit. In: Zeitschrift für Urheber- und Medienrecht, 61 (6), 461-476.

Wigge, P. (2000): Evidenz-basierte Richtlinien und Leitlinien – Qualitäts- oder Steuerungsinstrumente in der GKV? In: MedizinRecht, 18 (12), 574-585.

Wilbanks, J. T.; Topol, E. J. (2016): Stop the privatization of big data. In: Nature, 353 (7612), 345-348.

Williams, H. et al. (2015): Dynamic consent: a possible solution to improve patient confidence and trust in how electronic patient records are used in medical research. In: JMIR Medical Informatics, 3 (1), art. no.: e3. DOI: 10.2196/medinform.3525.

Wirth, J. (2015): Goldgrube Big Scientific Data. In: GIT Spezial BIOforum, 38 (2), 24-26.

Wright, D.; De Hert, P. (Hg.) (2016): Enforcing Privacy. Regulatory, Legal and Technological Approaches. Cham.

Zhelyazkova, A. et al. (2017): Munich Digital Healthcare Footprint. Food, Wellbeing & Trade. Ed. by Munich Digital Institute. https://www.munich-digital.com/download/article/?token=c461296644ad0420c35ceee5fd8229840d85ef9c [2017-12-09].

Zirfas, J. (2017): Smart Health rechtsverträglich gestaltet: Ubiquitous Computing in der Gesundheitspflege und -vorsorge. Wiesbaden.

Zook, M. et al. (2017): Ten simple rules for responsible big data research. In: PLoS Computational Biology, 13 (3), art. no.: e1005399. DOI: 10.1371/journal.pcbi.1005399.

CITED DECISIONS

BVerfG, 15 December 1983 – 1 BvR 209/83, 1 BvR 269/83, 1 BvR 362/83, 1 BvR 420/83, 1 BvR 440/83, 1 BvR 484/83 (BVerfGE 65, 1)

BVerfG, 11 March 2008 – 1 BvR 2074/05, 1 BvR 1254/07 (BVerfGE 120, 378)

BVerfG, 26 February 2008 – 1 BvR 1602/07, 1 BvR 1606/07, 1 BvR 1626/07 (BVerfGE 120, 180)

BVerfG, 11 June 1991 – 1 BvR 239/90 (BVerfGE 84, 192)

BVerfG, 17 July 2013 – 1 BvR 3167/08 (NJW 2013, 3086)

EuGH, 19 October 2016 – C-582/14 (NJW 2016, 3579)

BGH, 28 October 2014 – VI ZR 156/13 (NJW 2014, 1235)

BGH, 18 April 2013 – I ZR 53/09 (NJW-RR 2014, 46)

VG Darmstadt, 24 June 2004 – 1 E 470/04 (3) (NVwZ-RR 2006, 566)

ABBREVIATIONS

AA	Akademieausgabe (Academy edition)
Art.	Article
BDSG	Bundesdatenschutzgesetz (Federal Data Protection Act)
BGB	Bürgerliches Gesetzbuch (Civil Code)
BGBl.	Bundesgesetzblatt (Federal Law Gazette)
BGH	Bundesgerichtshof (Federal Court of Justice)
BVerfG	Bundesverfassungsgericht (Federal Constitutional Court)
BVerfGE	Entscheidungen des Bundesverfassungsgerichts (Decisions of the Federal Constitutional Court)
CDISC	Clinical Data Interchange Standards Consortium
CFR	Charter of Fundamental Rights of the European Union
ed.	edition/editor/edited
EDP	electronic data processing
EDPS	European Data Protection Supervisor
EGA	European Genome-phenome Archive
EuGH	Europäischer Gerichtshof (European Court of Justice)
EURODAC	European system for the comparison of fingerprints of asylum applicants
GDPR	General Data Protection Regulation
GenDG	Gendiagnostikgesetz (Genetic Diagnosis Act)
GG	Grundgesetz (Basic Law)
HTTP	Hypertext Transfer Protocol
ID	identification
IT	information technology
MPG	Medizinproduktegesetz (Medical Devices Act)
NJW	Neue Juristische Wochenschrift
NJW-RR	NJW-Rechtsprechungs-Report Zivilrecht
no.	number
NVwZ-RR	NVwZ-Rechtsprechungs-Report Verwaltungsrecht
OECD	Organization for Economic Co-operation and Development
OJ	Official Journal
para.	paragraph*
RFID	radio-frequency identification

SGB	Sozialgesetzbuch (Social Code)
SSO	single sign-on
StGB	Strafgesetzbuch (Criminal Code)
TFEU	Treaty on the Functioning of the European Union
VG	Verwaltungsgericht (Administrative Court)

* In the original German citations, “Rn.” (Randnummer, or literally “number on the margin”) refers to the practice of numbering paragraphs in the margin of many German legal opinions. Here, “Rn.” is translated as “para.”.

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