Big Data health research: Safeguarding rights and interests

Menno Mostert

Colofon

Big Data health research: Safeguarding rights and interests | Menno Mostert
PhD thesis. Department of Medical Humanities, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, the Netherlands.
ISBN: 978-94-028-1282-4

Cover image adapted from Andy Lamb | License: Attribution 4.0 International (CC BY 4.0)
Cover design by Elisa Calamita, persoonlijkproefschrift.nl
Layout and design by Elisa Calamita, persoonlijkproefschrift.nl
Printed by Ipskamp Printing, proefschriften.net

Big Data health research: Safeguarding rights and interests

Big Data in de medische wetenschap: beschermen van rechten en belangen
(met een samenvatting in het Nederlands)

PROEFSCHRIFT

Ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. H.R.B.M. Kummeling, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op

dinsdag 18 december 2018
des middags te 12.45 uur

door

Menno Mostert
Geboren op 2 februari 1984
te Utrecht

Promotoren:

Prof. dr. J.J.M. van Delden
Prof. dr. A.L. Bredenoord

Manuscripts based on the studies presented in this thesis

Chapter 2

Mostert M, Bredenoord AL, van der Sloot B, van Delden JJM. From privacy to data protection in the EU: Implications for Big Data health research. *European Journal of Health Law 2018; 25: 43-5.*

Chapter 3

Mostert M, Bredenoord AL, Biesaart MCIH, van Delden JJM. Big Data in medical research and EU data protection law: Challenges to the consent or anonymise approach. *European Journal of Human Genetics 2016; 24: 956-60.*

Chapter 4

Mostert M. Big data, medisch-wetenschappelijk onderzoek en gegevensbescherming. In: Ottes L., Kits PM, Zwenne GJ, Steenbruggen WAM, van Veen EB, Kleefman TG, Miedema F, Mostert M. *Big data in de zorg, preadvies uitgebracht voor de Vereniging voor Gezondheidsrecht, Sdu Uitgevers: Den Haag, 2017, p. 165-93.*

Chapter 5

Mostert M, Koomen BM, van Delden JJM, Bredenoord AL. Privacy in Big Data psychiatric and behavioural research: A multiple-case study. *International Journal of Law and Psychiatry 2018; 60: 40-4.*

Chapter 6

Kalkman S, Mostert M, van Thiel GJMW, van Delden JJM. Responsible data sharing in international health research: A review of principles and norms. *Submitted.*

Chapter 7

Bredenoord AL, Mostert M, Isasi R, Knoppers BM. Data sharing in stem cell translational science: Policy statement by the International Stem Cell Forum Ethics Working Party**.** *Regenerative Medicine 2015; 10: 857-61.*

Contents

Chapter 1

General introduction

Chapter 1

The rise of Big Data in health research

The term Big Data has often been used to describe the exponential growth of data. After Big Data was coined about two decades ago, it was suggested that the volume, velocity and variety of data (the three 'V's) can be considered as its defining qualities,[1] and later on additional 'V's have been proposed. Nowadays, the usefulness of the 'V's as defining qualities of Big Data is in dispute, and there is no concise or widely accepted definition of the term.[2]

In the context of health research, it is not only the exponential growth of data captured that is important to understand what Big Data is about. Mayer-Schönberger and Ingelsson emphasize that a combination of changes in data, methods and purpose of research activities resemble a shift towards Big Data analytics in health research.[2] They describe three key areas in which Big Data approaches differ from conventional data analyses in research (Fig. 1). Firstly, researchers more comprehensively capture data relative to phenomenon of interest. Consequently, they need to deal with a range of trade-offs, such as between data quality and data quantity. Secondly, researchers are building new data analysis methods to extract valuable information from the more comprehensive data. One of the most innovative approaches is the deployment of different forms of machine learning. Thirdly, they point out a shift in the purpose of data analyses. The purpose is not only to evaluate human generated hypotheses or answer existing questions. Big Data analytics is also deployed to recognise patterns in data that aid in hypothesis generation and raising relevant questions.

Irrespective of the exact definition of Big Data, what is needed for Big Data to become transformative in health research is the wide scale collection, reuse and linkage of data,[2] usually at the individual person level.[3] More and more infrastructures and initiatives exist to facilitate such reuse and linkage of data for a broad range of health research purposes, such as the UK Biobank and the Biobanking and BioMolecular resources Research Infrastructure-European Research Infrastructure Consortium (BBMRI-ERIC).[4,5] Another example is the BigData@Heart consortium, which applies Big Data approaches to improve patient outcomes related to the most common cardiovascular diseases.[6] In its efforts to achieve this aim, BigData@Heart is bringing together rich sets of health-related data on over 25 million individuals. These developments exemplify how health research is becoming a *data-intensive* activity, in which health-related, genomic and other data about individuals are captured, reused and linked on a massive scale.

Small data approach		Big Data approach
Optimised for data quality and cost	*Data*	Optimised for comprehensiveness and insight
Conventional statistics	*Typical methods*	Machine learning
Answering questions	*Purpose*	Also generating hypotheses

Figure 1. Where Big Data analytics differs from small data analytics, according to Mayer-Schönberger and Ingelsson, adopted from[2].

Normative challenges and legislative developments

Key normative concerns related to Big Data health research, that have been identified in a review of the literature, relate to privacy, data protection, informed consent, anonymisation, ownership, epistemology, (the myth of) objectivity, and Big Data divides or inequalities.[7] The key areas of concern *and* legislative activity in the EU have been privacy and data protection.

The need for a more consistent and comprehensive protection of personal data was recognised in the EU, and a reform of data protection law was initiated. The Data Protection Directive 95/46/EC were to be replaced by a regulation, which is a powerful instrument of the EU to adopt rules that are immediately binding in all Member States. The potential tension between new data protection requirements and interests in health research became apparent during this reform of data protection law, which took place simultaneously with the completion of this thesis. Many stakeholders feared that the regulation, as proposed by the EU Parliament, would severely restrict health research, mainly because of the combination of strict consent requirements and limited research exemptions.[8] Ultimately, the law reform in the EU resulted in the adoption of the General Data Protection Regulation (GDPR), which entered into force in May 2018.[9]

Although the final version of the GDPR is not as strict as initially feared,[10] it remains a challenge to interpret, balance, implement and harmonise its principles and rules related to health research. Despite the GDPR's direct effect in EU Member States, many provisions relevant to health research do need to be transposed into national law.[11] As a consequence, a high level of harmonisation was not achieved in the GDPR. Moreover, the GDPR is still largely based on the same principles as Directive 95/46/EC, many of which are at odds with Big Data approaches. The large-scale reuse and linkage of personal data seem difficult to reconcile with data protection

principles like purpose limitation, storage limitation and data minimisation. What is more, the discussion about the limits of informed consent and anonymity in safeguarding and balancing relevant interests remains equally important under the final GDPR. Another topic of discussion is related to a change in the key fundamental rights on which the GDPR is based, which could aid in a sound implementation and interpretation of its principles and rules in the context of health research. Article 1(2) of the GDPR affirms that it in particular protects the fundamental right to data protection. This is in contrast to the former Data Protection Directive 95/46/EC, which in particular protected the right to privacy with respect to the processing of personal data.

Against this background of normative complexity, change and debate, researchers and other stakeholders do engage with the challenge of utilising the potential of Big Data in health research. In the UK, a study has shown that the complex nature of the regulatory landscape resulted in a culture of caution and (overly) conservative approaches to data sharing.[12] Nevertheless, data reuse and sharing is increasingly regarded as a scientific and ethical imperative,[13,14] and the availability and use of data resources in health research is growing in a rapid pace.

Aim and scope

The main aim of this thesis is to inform the debate about what form laws, regulations and information governance should take in the EU, to allow for progress in data-intensive health research while safeguarding (fundamental) rights and morally relevant interests. To achieve this aim, this thesis addresses the central question of how relevant rights and interests can be safeguarded and balanced in the EU, without disproportionately hampering data-intensive health research. The central question is addressed through a number of sub-questions:

I. Are there differences between the right to data protection and the right to privacy in the EU, which are relevant in the context of data-intensive health research?
II. How is the consent or anonymise approach challenged in a data-intensive health research context, and what are possible ways forward within the EU legal framework on data protection?
III. Does the GDPR contribute to a responsible and effective use of personal data in data-intensive health research?

IV. What challenges related to privacy- and data protection are encountered in real-world examples of data-intensive health research?
V. What are the ethically relevant principles and norms so far developed by (international) working groups or professional organisations with respect to data sharing in health research?
VI. What are the specific policy principles for responsible data sharing in stem cell translational science?

Structure

In chapters 2, 3 and 4, the key challenges and ways forward in the EU legal framework on privacy and data protection, which are relevant to data-intensive health research, are discussed. Questions I, II and III are addressed in these chapters. Subsequently, chapters 5, 6 and 7 provide insight in other relevant sources of normativity, and show how concerns can be dealt with in specific contexts. Questions IV, V and VI are addressed in these chapters.

Chapter 2 reflects on two of the key fundamental rights that underpin EU data protection law; the right to privacy and the right to data protection. It shows that there are multiple differences between these two rights, which are relevant to data-intensive health research. **Chapter 3** reviews how the consent or anonymise approach is challenged in data-intensive health research, and discusses possible ways forward within the changing EU legal framework on data protection. **Chapter 4** examines whether the GDPR has achieved its dual objective of both facilitating health research and subjecting it to appropriate safeguards. Key issues in this analysis are anonymity, the consent requirement, and research exemptions related to consent, data processing principles and individual rights.

Chapter 5 presents the results of the multiple-case study. By means of this qualitative study, insight is gained in how privacy and data protection concerns are currently dealt with in two real world examples of Big Data health research. In **chapter 6**, the results of a review of ethical guidelines, policy documents and literature sources for ethical principles and norms pertaining to data sharing for international health research are presented. The aim is to identify a set of ethical principles and norms to govern responsible data sharing for international health research. **Chapter 7** focuses on data sharing in the specific context of stem cell science and discusses principles to provide guidance. These principles include engagement,

data quality and safety, privacy, security and confidentiality, risk–benefit analysis and sustainability.

Finally, **chapter 8** reflects on the main findings in this thesis and discusses ways forward.

Research approach

In this thesis, various research methodologies have been used. A combination of normative, literature review and qualitative case study research approaches has been deployed.

The normative research is primarily focussed on EU law that governs the use of data, and in particular on the GDPR. Incidentally, references to national laws are made. These bodies of law form a starting point for the doctrinal and normative (legal) analysis. Doctrinal legal research is a form of positive legal research and seeks to explain what the law is. The goal of doctrinal legal research is to identify, analyse and synthesise the content of law.[15] Doctrinal legal research is more than a literature review of secondary sources. The primary data consists of the sources of law. Primary research is the step of locating and analysing different sources of the law and then adding novel information to the known body of law.[15] Criticism to this approach is often directed to the internal perspective of doctrinal research. In this internal perspective, positive law is regarded as being in a relatively autonomous relation to the social, political and economic reality.[16] In normative legal research, a broader perspective is used. It focusses on what the law ought to be and is evaluative. Fundamental rights can be seen as cornerstones for deciding how interests should be balanced and establishing what the law ought to be.[16] Fundamental rights are rather an important source of arguments, when balancing rights and interests. However, fundamental rights are often not sufficiently well defined to decide specific cases, since it is impossible to give one uniform answer to what one legally ought to do.[17] This normative uncertainty can be seen as an essential element of normative legal scholarship. In this view, normative legal scholarship is a discipline of conflicting arguments.[16] In normative legal analyses, existing law can also be considered as providing empirical input on how to deal with conflicting arguments.[16] Conform this perspective, legislation, rules and case law are not only considered as binding statements of what the law is, but rather as a source of information about normative arguments.

Normative legal analyses are often entwined with other (normative) theories, like theories from ethics or social sciences.[16] In this thesis, the normative legal analyses are in particular entangled with practical ethics. Practical ethics allows normative claims to be made about a certain practice, by presenting sufficient, consistent and coherent reasoning for them.[18] When making normative claims about a certain practice, it is important to relate these claims to the phenomenon under study in a well-informed and proactive way. After all, technology, science, society, ethics and law should not be regarded as separate worlds. They inform and shape each other in a process that is captured by the notion of 'co-production'. Co-production refers to the simultaneous processes through which epistemic and normative understandings of the world are formed.[19] This underlines the importance of incorporating empirical data in normative analyses. To this end, literature has been studied and the need for a qualitative multiple-case study has been identified. The case study is a research methodology that allows the in depth evaluation of a certain phenomenon within its real-world context.[20,21] In the case study, an iterative and inductive research process was followed, in which the normative analysis proceeded during and alongside data collection and analysis.[22] This ensured an interchange between the normative and empirical, and allowed the refinement and narrowing of understanding and argumentation over time.

References

1. Laney D. 3D data management: controlling data volume, velocity and variety. META Gr Rep 2001; 6: 70.
2. Mayer-Schonberger V, Ingelsson E. Big Data and medicine: a big deal? J Intern Med 2018; 283: 418–29.
3. Weber GM, Mandl KD, Kohane IS. Finding the Missing Link for Big Biomedical Data. JAMA 2014; 311: 2479–80.
4. Allen NE, Sudlow C, Peakman T, Collins R, UK Biobank. UK Biobank Data: Come and Get It. Sci Transl Med 2014; 6: 224ed4-224ed4.
5. van Ommen G-JB, Törnwall O, Bréchot C et al. BBMRI-ERIC as a resource for pharmaceutical and life science industries: the development of biobank-based Expert Centres. Eur J Hum Genet 2015; 23: 893–900.
6. Hemingway H, Asselbergs FW, Danesh J et al. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. Eur Heart J 2017. doi:10.1093/eurheartj/ehx487.
7. Mittelstadt BD, Floridi L. The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. Sci Eng Ethics 2016; 22: 303–41.
8. Ploem MC, Essink-Bot ML, Stronks K. Proposed EU data protection regulation is a threat to medical research. BMJ 2013; 346: f3534.
9. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016. Available at https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:32016R0679.
10. Dove ES, Thompson B, Knoppers BM. A step forward for data protection and biomedical research. Lancet (London, England) 2016; 387: 1374–75.
11. The Wellcome Trust. Analysis: Research and the General Data Protection Regulation - 2012/0011(COD) July 2016 (v1.4). Available at https://wellcome.ac.uk/sites/default/files/new-data-protection-regulation-key-clauses-wellcome-jul16.pdf.
12. Sethi N, Laurie GT. Delivering proportionate governance in the era of eHealth. Med Law Int 2013; 13: 168–204.
13. Knoppers BM, Harris JR, Budin-Ljøsne I, Dove ES. A human rights approach to an international code of conduct for genomic and clinical data sharing. Hum Genet 2014; 133: 895–903.
14. Ohno-Machado L. To share or not to share: that is not the question. Sci Transl Med 2012; 4: 165cm15.
15. Watkins, D. (Ed.), Burton, M. (Ed.). (2018). Research Methods in Law. London: Routledge.
16. Smits JM, The Mind and Method of the Legal Academic, Cheltenham: Edward Elgar Publishing, 2012 (ISBN: 9780857936554).
17. Singer JW, Normative Methods for Lawyers, UCLA Law Review, 2009; 56: 922, available at http://www.uclalawreview.org/pdf/56-4-3.pdf.

18. Beauchamp TL, Childress J. Principles of Biomedical Ethics. 7th ed. Oxford University Press: New York, 2012.
19. Jasanoff S (ed.). States of Knowledge: The Co-Production of Science and the Social Order. Taylor & Francis: Abingdon, UK, 2004 doi:10.4324/9780203413845.
20. Yin, R. K. (2014). Getting started: How to know whether and when to use the case study as a research method. In R. K. Yin (Ed.), Case study research: Design and methods (pp. 3–25). CA, USA: Thousand Oaks.
21. Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. The Qualitative Report, 13, 544–59.
22. Dunn M, Sheehan M, Hope T, Parker M. Toward Methodological Innovation in Empirical Ethics Research. Cambridge Q Healthc Ethics 2012; 21: 466–80.

Chapter 2

From privacy to data protection in the EU: Implications for Big Data health research

Mostert M
Bredenoord AL
van der Sloot B
van Delden JJM

European Journal of Health Law 2018; 25: 43-5.

Abstract

The right to privacy has usually been considered as the most prominent fundamental right to protect in data-intensive (Big Data) health research. Within the European Union (EU), however, the right to data protection is gaining relevance as a separate fundamental right that should in particular be protected by data protection law. This paper discusses three differences between these two fundamental rights, which are relevant to data-intensive health research. Firstly, the rights based on the right to data protection are of a less context-sensitive nature and easier to enforce. Secondly, the positive obligation to protect personal data requires a more proactive approach by the EU and its Member States. Finally, it guarantees a more comprehensive system of personal data protection. In conclusion, we argue that a comprehensive system of data protection, including research-specific safeguards, is essential to compensate for the loss of individual control in data-intensive health research.

Introduction

Over the last decade, technical possibilities for collecting, re-using and linking data related to individuals have increased tremendously. Moreover, data sharing for health research purposes is increasingly being presented as an ethical and scientific imperative.[1] The effectiveness of certain traditional approaches that govern the use of data in health research is, however, decreasing in the era of Big Data. It has been indicated that a strict 'consent or anonymise approach' neither sufficiently allows for progress in data-intensive health research, nor adequately protects individual rights and interests.[2,3] In addition, the large scale re-use of data is difficult to reconcile with certain data protection principles, such as purpose limitation and data minimisation.[4] The current debate is about what form laws and information governance — consisting of organisational and technical measures — should take to allow for progress in data-intensive health research while effectively protecting fundamental rights and other morally relevant interests.

This debate usually revolves around the right to respect for private life (hereafter: the right to privacy) as the key fundamental right to protect. Within the EU, however, an independent fundamental right to data protection gradually emerged in addition to the right to privacy.[5] After its separate recognition in the EU Charter of Fundamental Rights, the right to data protection acquired a prominent position in the EU General Data Protection Regulation 2016/679 (GDPR), which will apply from 25 May 2018. Article 1(2) of the GDPR unambiguously affirms that it protects fundamental rights and in particular the right to data protection. This is in contrast to the current EU Data Protection Directive 95/46/EC, which protects in particular the right to privacy with respect to the processing of personal data. This change in emphasis is reflected throughout the whole GDPR and therefore also in provisions related to health research. Article 9(2)(i) of the GDPR is such a provision, which only allows the use of special categories of personal data in health research without consent, when the law provides a derogation that respects the essence of the right to data protection.

It largely remains unclear what this shift from the right to privacy to the right to data protection in the EU means. There is an ongoing debate about the differences between both rights and the rationale for introducing data protection as an independent right.[6,7] This uncertainty could negatively impact a coherent interpretation and implementation of both fundamental rights and the provisions in the GDPR relevant to data-intensive health research. The aim of this paper is to

clarify this matter by discussing whether there are differences between the right to data protection and the right to privacy relevant within the context of data-intensive health research.

A right to data protection in the Charter of Fundamental Rights of the EU

In the EU, a fundamental right to data protection sits alongside the right to privacy. The Charter of Fundamental Rights of the EU (the Charter) contains a right to the protection of personal data in Article 8 (the right to data protection), in addition to a right to respect for private life in Article 7 (the right to privacy). In 2009, legally binding force was granted to the Charter in the Lisbon Treaty and the Charter acquired the status of primary EU law. According to its preamble, the Charter "reaffirms" fundamental rights in the EU and makes them "more visible" to strengthen the protection of those rights. Some scholars, however, underline that the Charter did not reaffirm or make the right to data protection more visible, but actually created such a right in addition to the right to privacy.[5] Moreover, the impact of the right to data protection as a separate right is increasingly visible in case law of the Court of Justice of the EU (CJEU).[8,9] In addition, as mentioned above, Article 1(2) of the GDPR now clearly affirms that the Regulation 'protects fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data.' In addition, familiar terms, such as "privacy by design" and "privacy impact assessment" have been replaced in the GDPR by "data protection by design" and "data protection impact assessment" (Articles 25 and 35 of the GDPR). Nearly all other references to privacy or the private life have also disappeared in both the legislative text and the recitals.

This way of framing data protection norms in the Charter, the jurisprudence of the CJEU and the GDPR is different from how it has been framed in traditional data protection instruments and case law of the European Court of Human Rights (ECtHR). In the Organisation for Economic Co-operation and Development (OECD) context, national laws on data protection are typically referred to as 'privacy laws'.[5] In Convention 108 of the Council of Europe and EU Directive 95/46/EC, data protection norms are presented as serving in particular the right to privacy. Since the right to data protection, as such, is not included in the European Convention on Human Rights (ECHR), the competence of the ECtHR is limited to personal data processing activities that fall within the scope of Article 8 ECHR, or another right in the ECHR.

Personal data processing could fall within the scope of Article 8 ECHR, when the personal data processing engages aspects of the private life. Whether this is the case, depends on the nature of the data, the context in which the data is processed, the way the data is used and the results of the processing.[10]

Furthermore, it should be taken into consideration that the Charter, in itself, is different from traditional human rights instruments, such as the ECHR, in a complex way.[11] The Charter is not a freestanding bill of rights with a universal scope. According to Article 51 of the Charter, it applies to EU institutions and Member States only when they are implementing EU law. Nevertheless, EU and Member State law should, as a minimum, be in accordance with the Charter. Consequently, a provision in EU or Member State law could no longer be applicable when it is in conflict with the Charter.[12] An important function of the Charter, therefore, is to guide the implementation and interpretation of EU law, including the GDPR.

How data protection differs from privacy

At first glance, it seems like the right to data protection has dethroned the right to privacy as the key fundamental right to protect, according to Article 1(2) of the GDPR. A closer study however reveals that the reality is more complex, mainly because of the complicated relationship between both rights. In the Charter's explanatory memorandum, it is emphasized that the right to data protection is partially based on the right to privacy.[13] Unfortunately, the memorandum does not adequately explain the justification of a separate introduction of the right to data protection. In addition, there seems to be a large overlap between the scope of both rights.[14] Moreover, both rights serve many of the same objectives.[7] This, combined with the difficulties in defining the right to privacy, makes it difficult to draw a sharp distinction between the two rights. A growing number of legal scholars nevertheless agrees that the right to data protection should not be regarded as an element of, or a mere derivation from, the right to privacy. Moreover, they agree that relevant differences between both rights exist.[5,8,9,14,15] Below, we identify and discuss three of the differences between the right to data protection and the right to privacy, that we consider most relevant.

Individual rights decoupled from privacy

Firstly, both the scope and the substance of the individual rights guaranteed by the right to data protection differ from those based on the right to privacy. It is the mere processing of personal data that allows data subjects to invoke their rights based on the right to data protection. The definitions of 'personal data' and 'processing' are broad. According to Article 4 sub 1 and 2 of the GDPR, these terms cover any operation which is performed on any information relating to a natural person who can be identified, directly or indirectly. Consequently, almost all forms of personal data processing fall under the scope of the right to data protection, regardless of whether the right to privacy is interfered with.[8,15] In contrast, whether or not the right to privacy is interfered with depends on both the nature and the context of the specific processing.[7,9] This difference in scope is illustrated by some of the judgements of the CJEU. In the *Rundfunk* judgement, the Court held that "(…) the mere recording by an employer of data by name relating to the remuneration paid to his employees cannot as such constitute an interference with private life". According to the Court, the recording of personal data, by itself, thus did not fall within the scope of the right to privacy, whereas the Court noted that such a recording falls within the scope of the right to data protection since it constitutes personal data processing.[16] Furthermore, in the *Digital Rights Ireland* case, the CJEU confirmed that the retention of personal data also directly and specifically affects the right to privacy, when the "(…) data, taken as a whole, may allow very precise conclusions to be drawn concerning the private lives of the persons whose data has been retained, such as the habits of everyday life, permanent or temporary places of residence, daily or other movements, the activities carried out, the social relationships of those persons and the social environments frequented by them."[17] The individual rights based on the right to privacy are, therefore, of a more context-sensitive nature.

In addition to the difference in scope, the substantive protection offered by the right to privacy and the right to data protection also differs. This is illustrated by the confirmation of the ECtHR that the right to privacy does not guarantee a general right of access by the data subject to his own personal data.[18] This is in contrast to the right to data protection, which explicitly guarantees such a right of access in the abstract, irrespective of whether there is an interference with the right to privacy. Some, however, argue that the ECtHR is currently moving towards the introduction of a more general right of access, based on the right to privacy.[7] This growing willingness of the ECtHR to recognise more general rights, based on the right to privacy, makes it increasingly difficult to discern a distinction between the substantive protection

offered by both rights. Differences between the substantive protection offered by the right to data protection and the right to privacy do nevertheless remain.[7,8] These differences may be related to the dissimilar background of the right to data protection, which is also designed to protect non-privacy related interests.

A more positive approach

A second difference is that the right to data protection has been designed as a largely positive obligation of the EU and its Member States. To fulfil this positive obligation, governments will need to take affirmative measures to protect personal data. In addition, the right to data protection has been designed to regulate both horizontal and vertical relationships. This is in contrast to the right to privacy, which was originally coined as a mere negative obligation of public authorities to refrain from arbitrary interference with the private lives of individuals.[19] The ECtHR still considers this negative obligation as the essential object of the right to privacy.[20]

Today, positive obligations related to data-processing activities of private sector entities are nevertheless also inferred from the right to privacy. The ECtHR confirmed that states may be required to adopt measures designed to secure respect for the right to privacy, "even in the sphere of the relations of individuals between themselves".[21] These positive obligations based on the right to privacy do, however, suffer from a number of limitations. One of these limitations is that the concrete positive obligations stemming from the right to privacy are always linked to particular circumstances. This is because what constitutes these positive obligations is predominantly determined by the ECtHR on a case-by-case basis. These cases do not provide a basis for the more general positive obligations as guaranteed by the right to data protection.[8]

The right to data protection therefore complements the positive obligations inferred from the right to privacy with explicit positive obligations that are of a more abstract nature. Consequently, the somewhat blurred distinction between privacy as an essentially negative obligation and data protection as a largely positive obligation is still relevant.

A more comprehensive and systematic approach

A third difference is that the right to data protection rests on a more comprehensive and systematic approach, one beyond individual rights. Article 8 of the Charter guarantees a comprehensive system of data protection norms and explicitly confirms that the principles of fair and lawful processing, purpose specification and limitation,

and the requirement of independent supervision are key elements of this system. In addition, data security — consisting of technical and organisational measures to prevent the accidental loss, alteration or unlawful destruction of the data — was referred to by the CJEU as an essential element of the right to data protection.[17] Other key elements of EU data protection law, such as accountability and data quality requirements, may also implicitly be guaranteed by Article 8(1) of the Charter. Therefore, the right to data protection does not solely rely on individuals who exercise or enforce their rights, but is also based on a set of duties addressed to a broad range of actors involved in personal data processing. Although some of these duties may correlate with individual rights, this is not necessarily the case. An example is that compliance with data protection rules should be subject to control by an independent authority. A similar obligation, just as comprehensive, may not directly result from the case law of the ECtHR based on the right to privacy,[15] especially when it comes to the protection of individuals in horizontal relationships.

The extent to which the right to privacy could embrace similar data protection requirements however remains a complicated matter, since the recognition of data protection norms based on the right to privacy is on a case-by-case basis. Although data security is for instance not regarded as an essential element of the right to privacy,[15] a lack of security measures could result in a violation of the right to privacy, especially when it concerns sensitive health information.[22] Nevertheless, the right to privacy is not considered to be of a nature to include independent supervision, data security or data quality requirements as its core elements. In other words, the right to privacy does not guarantee a comprehensive system of data protection norms similar to that guaranteed by Article 8 of the Charter.

Relevance to data-intensive health research

In the coming years, the EU and its Member States will need to fulfil their positive obligations based on the right to data protection, which have partially been encoded in the GDPR. Moreover, both public authorities and private sector entities will need to interpret the GDPR in accordance with fundamental rights. The increased emphasis on the right to data protection in the GDPR does, however, not necessarily render the right to privacy less relevant, especially in the context of data-intensive health research. After all, health research usually involves the processing of special categories of personal data, such as data concerning health or genetic data, which

often engages sensitive aspects of the private life. The right to data protection nevertheless adds an important layer of protection, as we discuss below.

The impact of individual rights

The individual rights rooted in Article 8 of the Charter could have a significant impact on data-intensive health research. Even though the right to data protection guarantees a system of data protection beyond individual rights, the individual rights of data subjects are still an essential element of this system. This may be why the allowed derogations from some of the individual rights in the GDPR are of a limited nature, especially when these rights are guaranteed by Article 8 of the Charter. Derogations from the right of access and the right to rectification (Article 8(2) of the Charter) for scientific research purposes may only be provided by law 'in so far as the individual rights would render impossible or seriously impair the achievement of the specific purposes(..)' (Article 89(2) in conjunction with Articles 15 and 16 of the GDPR). Moreover, derogations or exceptions from the right to information are not allowed at all when personal data are collected from the data subject himself (Article 13 of the GDPR). This right to information of the data subject is part of what constitutes "fair" processing, as referred to in Article 8(2) of the Charter.

A negative impact of these individual rights on data-intensive health research may nevertheless be reduced by taking them into account throughout the process of engineering information systems and shaping information governance. Those responsible for Big Data infrastructures and projects know beforehand which rights data subjects could invoke. This is due to the decoupling of the scope of individual rights of data subjects from an interference with the right to privacy, which results in more legal certainty. Implementing technical and organisation measures to ensure that data subjects can invoke their rights and that data-protection principles are implemented is not a mere opportunity for data controllers. It also is a legal obligation laid down in Article 25 of the GDPR under the title "Data protection by design".

Safeguards beyond individual rights and consent

The more positive and comprehensive approach required by the right to data protection is of great importance to allow progress in data-intensive health research in a responsible way. The key strength of the system of data protection is that it does not merely rely on strengthening individual rights or consent requirements to protect and balance relevant rights and interests.

After all, individuals are often no longer able to make meaningful decisions about the use of their personal data, as a consequence of the rapidly increasing scale and complexity of data-intensive health research.[2] Although efforts are made to enhance the exercise of individual control in health research by the use of online portals and engaging individuals as active participants,[23] it must be recognised that individuals can only selectively choose to be engaged. 'Broad consent' models, as referred to in Recital 33 of the GDPR, do recognise this to some extent by inviting people to agree to a broad range of future data use in research. This however inevitably leads to a trade-off between obtaining consent in a simple and practicable way, and providing individuals with sufficient information and control. Moreover, strengthening individual rights and consent requirements does not necessarily, in itself, reduce the risks to which individuals are exposed. What is more, merely relying upon consent and individual rights would not only result in an ineffective protection of individuals and their interests, it could also disproportionately hamper progress in data-intensive health research.[24] This is because it is often impracticable or impossible to allow individuals to exercise meaningful control over the use of their personal data in data-intensive health research.

The EU legislative bodies seem to have taken these considerations into account, not only by allowing derogations in favour of scientific research from consent requirements and some of the individual rights,[25] but also by requiring that such derogations should be subject to appropriate safeguards in accordance with the GDPR and the rights and freedoms of the data subject (Article 89(1) GDPR). In addition, when derogating from the obligation to obtain consent for the use of special categories of personal data for scientific research purposes, Article 9(2i) of the GDPR explicitly underlines the importance of respecting the essence of the right to data protection and providing for suitable and specific safeguards by law. By means of these derogations, the EU aims to facilitate scientific research, as long as the processing of personal data is subject to appropriate conditions and safeguards set out in EU or Member State law (Recital 157 GDPR). An important part of these derogations and safeguards, however, still need to be implemented in Member State law.[26] It thus becomes clear that respecting the right to data protection, while sufficiently allowing for progress in data-intensive health research, requires proactive legislators. When the EU and its Member States take this positive obligation serious, the GDPR could indeed be regarded as a step forward for data protection and health research.[27]

By way of contrast, the effectiveness of data protection law in regulating data-intensive health research has also been criticised. Some scholars have argued that

the term personal data is poorly defined and have raised questions about what data or communications should be protected by law.[28] Others have suggested that the limits of the law should be recognised and the strengths of soft law options such as ethical guidance or professional codes should be more appreciated.[29] In their view, data protection law should provide for sufficiently open norms to allow for soft law instruments, such as the international governance frameworks that are currently being developed.[30] The GDPR seems to meet this requirement, since Article 89(1) of the GDPR does not impose any strict safeguards on personal data processing for scientific research purposes. According to Article 89(1) of the GDPR, appropriate safeguards should "ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation". This obligation is weakened by adding that measures of data minimisation, which may include pseudonymisation, only need to be taken when the research purposes can be fulfilled in that manner. Moreover, Article 5(1c) of the GDPR already requires similar measures of data minimisation as an overarching safeguard.

Nonetheless, Article 89(1) of the GDPR does play a pivotal role in the protection of personal data when derogations from consent or individual rights are provided in favour of health research. In addition, as long as the data processing is in accordance with this provision, the re-use of personal data for scientific research purposes is not considered to be incompatible with the principles of purpose limitation and data minimisation (Article 5(1b) of the GDPR). It is therefore striking that Article 89(1) of the GDPR only provides very limited points of departure for what specific safeguards should be in place in a research context.

Conclusion

Although the rights to privacy and data protection are closely related, they should not be considered as identical. The right to data protection adds a crucial layer of protection beyond essentially negative obligations, individual rights based on the right to privacy, and consent requirements. It aims to complement the right to privacy by positively guaranteeing a more comprehensive and harmonised system of data protection norms, which are relatively easy to enforce and comply with.

Within the context of data-intensive health research, such a comprehensive system of data protection should be considered to serve two functions in particular. Firstly, the aim is to provide effective *overarching safeguards* that secure the rights and interests of individuals, irrespective of whether the personal data processing

is grounded on consent or any other legal basis. After all, merely adhering to the principle of lawfulness is never sufficient to respect the right to data protection. Secondly, such a system of data protection arranges for *specific safeguards* when it is necessary and proportional to derogate from consent requirements or certain individual rights. These specific safeguards are also essential to allow for the re-use of personal data in data-intensive health research, without taking heed of the principle of purpose limitation. The overarching safeguards should, amongst other things, include requirements of accountability subject to independent oversight, transparency towards data subjects and the public, ensure that data subjects can invoke their rights and data security. The issue of which specific safeguards should be provided for by law with regard to data-intensive health research remains unclear and deserves further study. After all, these specific safeguards should compensate for the loss of individual control as a result of the exceptions from individual rights and consent requirements for health research purposes.

At the same time, the limits of data protection law should be recognised. Relying on the distinction between personal and non-personal data to protect privacy and other relevant rights and interests might prove to be inadequate. In addition, inflexible or static data protection laws could hamper the development of suitable information governance frameworks on the national or international scale, in which the myriad of ethical, legal, social and professional norms need to be reconciled.

References

1. Knoppers BM, Harris JR, Budin-Ljøsne I, Dove ES. A human rights approach to an international code of conduct for genomic and clinical data sharing, Human Genetics 2014; 133: 895-903.
2. Mostert M, Bredenoord AL, Biesaart MCIH, van Delden JJM, Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach, European Journal of Human Genetics 2016; 24: 956-60.
3. Nuffield Council on Bioethics, The collection, linking and use of data in biomedical research and health care: ethical issues, 2015. Available at http://nuffieldbioethics.org/project/biological-health-data/, retrieved 20 January 2017.
4. Custers B, Uršič H, Big data and data reuse: a taxonomy of data reuse for balancing big data benefits and personal data protection, International Data Privacy Law 2016; 6: 4-15.
5. Fuster GG. The Emergence of Personal Data Protection as a Fundamental Right of the EU, Springer: Dordrecht, 2014.
6. Tzanou M. Data protection as a fundamental right next to privacy? 'Reconstructing' a not so new right, International Data Privacy Law 2013; 3: 88-99.
7. Lynskey O. Deconstructing data protection: the 'Added-value' of a right to data protection in the EU legal order, International and Comparative Law Quarterly 2014; 63: 569-97.
8. Hustinx P. EU Data Protection Law: The Review of Directive 95/46/EC and the Proposed General Data Protection Regulation, 2014. Available at https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/EDPS/Publications/Speeches/2014/14-09-15_Article_EUI_EN.pdf.
9. Kokott J, Sobotta C. The distinction between privacy and data protection in the jurisprudence of the CJEU and the ECtHR, International Data Privacy Law 2013; 3: 222-8.
10. ECtHR, Khelili v. Switzerland, App no. 16188/07 (18 October 2011).
11. S. Douglas-Scott. 'The European Union and Human Rights after the Treaty of Lisbon', Human Rights Law Review 11(4) (2011) 645-682.'
12. CJEU, Case C-399/11, Melloni, ECLI:EU:C:2013:107.
13. Convention Praesidium. Explanations Relating to the Charter of Fundamental Rights of the European Union, Brussels', 11 October 2000, Charte 4473/00, Convent 49.
14. Gellert R, Gutwirth S. The legal construction of privacy and data protection, Computer Law & Security Review 2013; 29: 522-30.
15. de Hert P, Gutwirth S. Data Protection in the Case Law of Strasbourg and Luxembourg: Constitutionalism in Action, in: S. Gutwirth et al. (eds.), Reinventing Data Protection? Springer: New York, 2009, pp. 3-43.
16. CJEU, Case C-139/01, Österreichischer Rundfunk and Others, ECLI:EU:C:2003:294, para. 74 and 64.
17. CJEU, Case C-293/12, Digital Rights Ireland, ECLI:EU:C:2014:238.

18. ECtHR, Gaskin v. United Kingdom, App no. 10454/83 (7 July 1989).
19. van der Sloot B. Privacy as human flourishing: Could a shift towards virtue ethics strengthen privacy protection in the age of Big Data?, Journal of Intellectual Property, Information Technology and Electronic Commerce Law 2014; 5: 230-44.
20. ECtHR, Hämäläinen v. Finland, App no. 37359/09 (16 July 2014).
21. ECtHR, X and Y v. the Netherlands, App no. 8978/80 (26 March 1985).
22. I. v. Finland, App no. 20511/03 (17 July 2008).
23. Kaye J, Whitley EA, Lund D, Morrison M, Teare H, Melham K. Dynamic consent: a patient interface for twenty-first century research networks, European Journal of Human Genetics 2015; 23: 141-6.
24. Di Lorio CT, Carinci F, Oderkirk J. Health research and systems' governance are at risk: should the right to data protection override health?, Journal of Medical Ethics 2014; 40: 488-92.
25. For an overview of these derogations see: The Wellcome Trust, Analysis: Research and the General Data Protection Regulation, 2016. Available at https://wellcome.ac.uk/sites/default/files/new-data-protection-regulation-key-clauses-wellcome-jul16.pdf.
26. The Wellcome Trust, Implementing the General Data Protection Regulation [2016/679] to maintain a competitive environment for research in Europe, 2016. Available at http://www.scienceeurope.org/wp-content/uploads/2016/10/EU-GDPR-implementation-Sep-2016.pdf.
27. Dove ES, Thompson B, Knoppers BM, A step forward for data protection and biomedical research, The Lancet 2016; 387: 1374-75.
28. O'Neill O. Can Data Protection Secure Personal Privacy?, in: T.S. Kaan, C.W. Ho (eds.), Genetic Privacy, 2013. Imperial College Press: London, pp. 25-40.
29. Laurie GT, Sethi N. Delivering proportionate governance in the era of eHealth: Making linkage and privacy work together, Medical Law International 2013; 13: 168-204.
30. Knoppers B.M. Framework for responsible sharing of genomic and health-related data, The Hugo Journal 2014; 8: 3.

Chapter 3

Big Data in medical research and EU data protection law: Challenges to the consent or anonymise approach

Mostert M
Bredenoord AL
Biesaart MCIH
van Delden JJM

European Journal of Human Genetics 2016; 24: 956-6

Abstract

Medical research is increasingly becoming data-intensive; sensitive data are being re-used, linked and analysed on an unprecedented scale. The current EU data protection law reform has led to an intense debate about its potential effect on this processing of data in medical research. To contribute to this evolving debate, this paper reviews how the dominant 'consent or anonymise approach' is challenged in a data-intensive medical research context, and discusses possible ways forwards within the EU legal framework on data protection. A large part of the debate in literature focuses on the acceptability of adapting consent or anonymisation mechanisms to overcome the challenges within these approaches. We however believe that the search for ways forward within the consent or anonymise paradigm will become increasingly difficult. Therefore, we underline the necessity of an appropriate research exemption from consent for the use of sensitive personal data in medical research to take account of all legitimate interests. The appropriate conditions of such a research exemption are however subject to debate, and we expect that there will be minimal harmonisation of these conditions in the forthcoming EU Data Protection Regulation. Further deliberation is required to determine when a shift away from consent as a legal basis is necessary and proportional in a data-intensive medical research context, and what safeguards should be put in place when such a research exemption from consent is provided.

Introduction

In recent years, both medical research and the legal landscape have been changing as a result of the rapid developments in information technology (IT). Medical researchers are collecting, re-using and linking health-related and genomic data on an unprecedented scale, based on the presupposition that this research will significantly improve human health.[1,2] Developments in IT have however led to an increasing concern about the effectiveness of existing data protection law, and the need for a more consistent and comprehensive protection of personal data was recognised in the European Union (EU).[3] Therefore, the Data Protection Directive 95/46/EC (DPD) is intended to be replaced by the General Data Protection Regulation (GDPR), which will be directly binding in all EU member states. On 12 March 2014, the European Parliament voted in favour of an amended draft GDPR.[4] The Council of the EU agreed on a common approach on a revised text of the proposed GDPR on 15 June 2015.[5] The final GDPR text depends on the outcome of the three-way negotiations between the Council, the Parliament and the European Commission. The ambition of the EU legislative bodies is to adopt the GDPR at the end of 2015.[6] After adoption, the GDPR will come into force after a transition period of likely 2 years.

The EU data protection law reform has led to an intense debate about its potential effect on medical research. Essentially, the discussion is about where limits should be drawn to the use of sensitive personal data in medical research. Resolving this matter requires a subtle negotiation of a broad range of relevant (fundamental) rights and interests. Key issues are related to the scope and limitations of consent as a legal basis for the use of sensitive personal data in medical research and its possible alternatives. A dominant approach in some EU member states is that the conventional or only alternative to obtain consent is anonymising these data. This has been referred to as the 'consent or anonymise approach'.[7,8] Even so, derogations to this approach can be laid down in data protection law in so-called 'research exemptions'.[9] This regulatory approach will continue to exist in the forthcoming GDPR, subject to still to be determined change in emphasis and detail. Both in literature and in the medical research community, many have expressed their concern about the consequences of the legislative reform. They indicate that the combination of strict consent requirements and limited research exemptions will severely restrict medical research.[10,11,12,13,14,15,16] To contribute to this evolving debate, this paper reviews how the consent or anonymise approach is challenged in a data-intensive medical research context, and discusses possible ways forward within the EU legal framework on data protection.

The context of data-intensive medical research

Increasingly large worlds of complex health-related and genomic data, often referred to as 'Big Data', are becoming available to medical researchers.[1] Initially, it was indicated that certain data characteristics define Big Data, like its relatively high volume, velocity and variety.[17] At present, the term is more and more used to refer to the technical or analytical methods to extract information from complex or multiple data sets.[1,18] Big Data sources potentially valuable to medical researchers include electronic health records (EHRs),[19] aggregated clinical trial data, administrative health care,[20] and genomic and other -omics data.[1,21] Nowadays, online activities of individuals, for example on mobile phones,[22] also allow the continuous collection of health-related and other data.[23] In the meanwhile, the wide-scale sharing of data is progressively promoted, for example in open access policies.[24] Furthermore, it is pointed out that linkage of multiple data sets at the individual person level is needed for Big Data to become transformative.[25]

Vital to the collection, re-use and linkage of multiple data sources on a large scale are the research infrastructures and networks in and outside the EU. For example, the UK Biobank provides access to medical researchers from all around the world to a wide variety of health-related data and human samples from more than 500 000 participants.[26] In Europe, the Biobanking and BioMolecular resources Research Infrastructure-European Research Infrastructure Consortium (BBMRI-ERIC) aims to facilitate the re-use of human samples and health-related data available in biobanks scattered across different nations.[27] Also, many initiatives exist to promote or facilitate the large-scale re-use and linkage of health-related and genomic data, such as the Global Alliance for Genomics and Health (Global Alliance).[28] These developments illustrate how medical research is increasingly becoming a data-intensive activity, in which health-related, genomic and other data are being collected, re-used and linked on a large scale.

EU legal framework on data protection

In the EU, the right to data protection and the right to privacy are formalised by an overlapping but different set of rules. This is because data protection law does not codify the right to privacy as such, but regulates the use of personal data, which are data related to identifiable individuals.[29] The right to data protection has recently been recognised as a separate fundamental right in Article 8 of the

Charter of Fundamental Rights of the EU (the Charter). Like any fundamental right, the right to data protection is not absolute and needs to be considered in its relation to other (fundamental) rights and interests, including the social rights of access to health care, social security and social assistance in case of illness (Articles 34 and 35 of the Charter), and the fundamental freedom of the sciences (Article 13 of the Charter). To this end, EU data protection law essentially provides a system of checks and balances, consisting of a set of principles and rules. At the heart of the current principal EU data protection law, the DPD, are the principles of fairness and lawfulness. The principle of fairness requires for example that those who process personal data are clear and open with individuals about how their data will be used. The principle of lawfulness demands that each processing of personal data must be based on consent or another legitimate basis laid down by law, as is also enshrined in Article 8 of the Charter.

When it comes to the processing of sensitive personal data, such as health-related data, a more restrictive set of legal bases is provided by EU data protection law. Genetic data will be explicitly recognised as sensitive in the forthcoming GDPR, without granting this type of data a status different from other categories of sensitive personal data.[30] At present, the legal base provided by Article 8 (2)(a) of the DPD for the processing of sensitive personal data, in any context, is explicit consent. For consent to be valid, it also needs to be specific, freely given and informed (Article 2 (h) DPD). Research exemptions from these consent requirements can be laid down in national law for reasons of 'substantial public interest', subject to the provision of 'suitable safeguards', according to Article 8 (4) of the DPD. Recital 34, which is related to this article, explicitly mentions that reasons of public interest can relate to areas such as scientific research and public health. It is however indicated that the implementation of research exemptions within national laws varies significantly between EU member states, and consequently hinders international collaboration between researchers.[31]

Ways forward within the consent or anonymise paradigm

Both the mechanism of consent and its conventional alternative of anonymisation are challenged in a data-intensive medical research context. Much of the debate, as outlined below, focuses on the legal or ethical acceptability of adapting consent or anonymisation mechanisms to overcome these challenges.

Adapting consent

The difficulties in obtaining consent, when personal data are to be available for linkage, re-use and analysis for largely undetermined future research purposes, have been discussed extensively in the literature.[32,33] On the one hand, it is questioned whether meaningful or legally valid (specific, explicit, freely given and informed) consent can be obtained at a one-off event at the time of data collection, as it may not be possible to foresee or comprehend the possible consequences of consenting.[9,34,35] On the other hand, it is suggested that obtaining specific consent for every linkage or re-use may be overly burdensome or impossible, because this could result in costly and time-consuming procedures, poor recruitment, consent bias, or unwarranted intrusion into the private lives of individuals.[36,37,38]

As a response to the difficulties in obtaining specific consent, adapted models of consent have been put into practice and discussed in the literature. The most common adaptations of consent are models that shift away from specific consent, such as 'broad consent', covering a broad range of future data uses.[32,33] There is however an ongoing debate on the legal validity and ethical acceptability of broad consent.[34,39,40,41,42] Some suggest that justifications for broad consent models remain contested in the bioethical literature, and they emphasise that these models are insufficient to ensure meaningful individual control over personal data or human samples.[9,43] Also, it is indicated that, effectively, broad consent is 'consent for governance' by certain institutions.[41] Others argue that broad consent is an ethically sound alternative for specific consent, although individuals are not given specific information about future research projects.[36,44]

In the draft GDPR texts, the current conflicting positions of the Parliament and Council on this topic appear to be reflected. Some indicate that broad consent may not meet the conditions on consent as defined in the Parliament's draft GDPR, regarding the information that must be given to the individual.[37,45] The position of the Council seems to be that broad consent should be possible for medical research.[16] This position is reflected in Recital 25aa of the Council's draft GDPR, which states that 'data subjects can give their consent to certain areas of scientific research when in keeping with recognised ethical standards for scientific research.' Moreover, Article 5 (1)(b) of the Council's draft GDPR provides a research exemption to the principle of purpose limitation, when appropriate safeguards are in place in accordance with Article 83.

An approach to consent claimed to be potentially consistent with strict or changing legal requirements is 'dynamic consent'. Essentially, dynamic consent

focuses on using IT and engaging individuals as active participants, so that they can be informed and subsequently re-consent can be obtained more easily.[46,47] Critics, however, argue that dynamic consent could for example lead to an information overload for the individual.[36] As a response to this critique, it is emphasised that dynamic consent is not a replacement for existing consent models, but rather a tool that could better facilitate the process of obtaining any form of consent.[47,48]

Adapting anonymisation mechanisms

A conventional method to protect data and avoid consent or other legal requirements is anonymisation. Yet, there seems to be a broad consensus that it is impossible to guarantee anonymity, especially when health-related data are re-used in different contexts or genomic data are involved.[8,49,50,51,52,53] Such a guarantee of absolute anonymity is however not required by data protection law. The term anonymisation is defined in current EU legal documents as a technique, which irreversibly prevents identification, taking into account all the means 'likely reasonably' to be used.[54] According to Recital 23 of the Parliament's draft GDPR, 'all the means reasonably likely to be used either by the controller or by any other person to identify or single out the individual directly or indirectly' should be taken into account in this assessment. In the Council's draft GDPR text, the phrase 'single out' has been removed from this recital.

Yet, it is indicated that irreversible anonymisation implicates extensive stripping of data sets, and largely excludes data linkage and update, while these activities are essential to most large research networks or projects.[55,56,57] Some therefore argue that lowering the thresholds for anonymisation could better balance relevant interests, by considering two-way coded data as de-identified in data protection law.[58,59] However, a more broadly accepted function of pseudonymisation (single or two-way coding) is considering it as a useful security measure.[54,60] In addition, Recital 23a of the Council's draft GDPR mentions that pseudonymisation can reduce risks, but is not intended to preclude the applicability of data protection law. It should be noted, however, that it is not the practical reality that a clear distinction between pseudonymous and anonymous data can always be made.[61] Another position is that anonymisation should be avoided in practice.[50,55] Not only since anonymisation excludes data linkage or update, but also because anonymisation takes away most legal obligations to protect the data or respect individual rights or interests, while the (hypothetical) risk of re-identification remains.[62] In addition, information

derived from anonymised data could still affect groups; risks of discrimination or stigmatisation have been described in the literature.[33,63]

The search for solutions with the use of anonymisation techniques and other innovative methods also carries on. An example is to prevent re-identification by 'taking the analysis to the data, not the data to the analysis', as facilitated by the initiative called dataSHIELD.[64] It is claimed that under DataSHIELD personal data re-use, linkage and analysis is enabled in accordance with legislation and guidance in the United Kingdom, primarily because no identifying or sensitive information is returned to the researcher.[65,66,67] Significant challenges however need to be overcome in the implementation of this initiative.[64]

Ways forward outside the consent or anonymise paradigm

An alternative approach is to search for ways forward outside the consent or anonymise paradigm, by creating another legal basis than consent for the processing of sensitive personal data for medical research purposes. According to Article 81 (2)(a) of the Parliament's draft GDPR, such a research exemption from consent should be provided by national law, for 'research that serves a high public interest'. In contrast, Article 9(2)(i) of the Council's draft GDPR indicates that consent is not required when the processing is necessary for scientific purposes, subject to certain conditions and safeguards laid down in law. Differing positions on the appropriate scope of research exemptions are also reflected in the literature. Some argue that research exemptions should be kept to a minimum by using dynamic consent approaches, taking into account the requirements of necessity and proportionality.[68] Others suggest that consent should serve as 'a default starting point from which departure is possible' for a particular data usage, when there is evidence of a strong justification in the public interest.[6] A more radical view is that providing another legal basis than consent should not be considered as an 'exemption', but as an equally acceptable route to achieve protection when data are re-used in large biobanks and data sets.[9] Also, some argue to reduce or eliminate the need for consent by focusing on solidarity arguments and harm mitigation.[69]

An interrelated issue is which appropriate safeguards should be put in place when a research exemption from consent is provided. In Article 81 of the Parliament's draft GDPR, mandatory pseudonymisation under the highest technical standards is presented as such a safeguard. It is argued though that a strict interpretation of this requirement will possibly render most data useless for epidemiological

research.[14] According to Article 83 (2) of the Council's draft GDPR, technological and/or organisational protection measures, such as pseudonymisation, could ensure that the processing of personal data is minimised, in pursuance of the proportionality and necessity principles. In addition, it does provide an escape where these measures would prevent achieving the scientific purpose, and this purpose cannot be fulfilled otherwise with reasonable means. Technological and organisational or governance measures have also been proposed in the literature to justify alternative legal bases to consent, such as opt-out registration,[9] authorisation by an ethics committee,[8] limiting data access and use, and engaging in public participation.[32] To overcome some of the challenges related to implementing governance mechanisms on an international scale, an e-governance system is proposed.[70]

Discussion

What can we learn from the above? In the debate on how to deal with the challenges to the consent or anonymise approach in the context of data-intensive medical research, within the EU legal framework on data protection, we suggest that the following considerations should be taken into account.

To begin with, we conclude that the search for ways forward within the consent or anonymise paradigm becomes increasingly difficult in a data-intensive medical research context. Although innovative technologies or methods could reduce some of these difficulties, a common position in the reviewed literature is that obtaining meaningful consent or irreversibly anonymising data is impracticable or impossible for a great deal of data-intensive medical research. It may be for these reasons that the necessity of a research exemption, which creates an alternative legal basis to consent, seems to be beyond questioning in the legal debate. This necessity may increase even further, dependent on what definitions on consent and anonymisation will be provided by the forthcoming GDPR, which need to be clear to reduce legal uncertainty and prevent the erosion of data protection law.

Then, we recommend that further debate should focus on two issues related to research exemptions in data protection law. First, we do not expect that a high level of harmonisation on the conditions of research exemptions will be provided by the forthcoming GDPR. The draft GDPR texts do provide an overlapping EU legal framework on this topic, but leave considerable room for a more detailed regulation on a national level. It therefore seems that it will be largely up to the EU member states to determine the appropriate conditions of research exemptions. This will

probably again result in a diverse implementation of research exemptions within the EU, which may impede the exchange of sensitive personal data for research across national borders. Initiatives within the medical research community to coordinate the development of harmonised approaches, such as BBMRI-ERIC and the Global Alliance, may therefore remain of vital importance to achieve the goal of international interoperability. Second, we notice that there is a lack of consensus on what the conditions of a research exemption from consent should be, while these conditions are of great influence to how relevant rights and interests need to be taken into account in a data-intensive medical research context. We agree with the suggestions in the literature that this act of balancing should include an independent necessity and proportionality test, for instance by an (data access) ethics committee. In addition, we emphasise that proportionate technical and governance measures should be incorporated in the design of data-intensive medical research projects and infrastructures, not only in order to provide a secure data processing environment, but also to allow individuals and the public to access clear information about the use of their data and their rights concerning this usage. Such transparency measures are in particular relevant where technological complexity makes it difficult for individuals to find out which personal data are used, for what purpose and by whom, as indicated in Recital 46 of both draft GDPR texts. We suggest that these measures could include the use of IT and participant interfaces to provide individuals with sufficient information and control over their data, and to stimulate participation by relevant stakeholders. Such a focus on research exemptions with appropriate safeguards should be preferred above continuing the practice of (over)stretching concepts of consent or anonymisation in order to sustain their central role. This may be necessary not only to meet legal requirements, but also to maintain public trust.

Overall, we conclude that research exemptions in data protection law should allow for the creation of a context-specific normative framework, in which the particularities of the use of sensitive personal data in medical research can be taken into account. Further interdisciplinary research is however needed to determine when a shift away from consent as a legal basis is necessary and proportionate in a data-intensive medical research context, and what technological and governance measures should be put in place when such a research exemption from consent is provided.

References

1. Costa FF. Big Data in biomedicine. Drug Discov Today 2014; 19: 433–40.
2. Mooney SJ, Westreich DJ, El-Sayed AM. Epidemiology in the era of Big Data. Epidemiology 2015; 26: 390–94.
3. Hustinx P. EU Data Protection Law: The Review of Directive 95/46/EC and the Proposed General Data Protection Regulation, 2014. Available at https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/EDPS/Publications/Speeches/2014/14-09-15_Article_EUI_EN.pdf.
4. European Parliament Legislative Resolution on the Proposal for a General Data Protection Regulation, 12 March 2014. Available at http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P7-TA-2014-0212+0+DOC+XML+V0//EN.
5. The General Approach of the Council of the EU on the General Data Protection Regulation, 15 June 2015. Available at http://data.consilium.europa.eu/doc/document/ST-9565-2015-INIT/en/pdf.
6. Council of the EU. Data Protection: Council Agrees on a General Approach, 2015. Available at http://www.consilium.europa.eu/en/press/press-releases/2015/06/15-jha-data-protection/.
7. Academy of Medical Sciences. Personal Data for Public Good: Using Health Information in Medical Research, 2006. Available at http://www.acmedsci.ac.uk/download.php?f=file&i=13206.
8. Sethi N, Laurie G. Delivering proportionate governance in the era of eHealth: making linkage and privacy work together. Med Law Int 2013; 13: 168–204.
9. Ruyter KW, LOuk K, Jorqui M, Kvalheim V, Cekanauskaite A, Townend D: From research exemption to research norm: recognising an alternative to consent for large scale biobank research. Med Law Int 2010; 10: 287–313.
10. Ploem MC, Essink-Bot ML, Stronks K. Proposed EU data protection regulation is a threat to medical research. BMJ 2013; 346: f3534.
11. Di Lorio CT, Carinci F, Oderkirk J. Health research and systems' governance are at risk: should the right to data protection override health? J Med Ethics 2014; 40: 488–92.
12. Kerr DJ. Policy: EU data protection regulation-harming cancer research. Nat Rev Clin Oncol 2014; 11: 563–64.
13. Dolgin E. New data protection rules could harm research, science groups say. Nat Med 2014; 20: 224.
14. Nyrén O, Stenbeck M, Grönberg H. The European Parliament proposal for the new EU General Data Protection Regulation may severely restrict European epidemiological research. Eur J Epidemiol 2014; 29: 227–30.
15. Dove ES, Townend D, Knoppers BM. Data protection and consent to biomedical research: a step forward? Lancet 2014; 384: 855.

16. Coppen R, van Veen EB, Groenewegen PP et al. Will the trilogue on the EU Data Protection Regulation recognise the importance of health research? Eur J Public Health 2015; 25: 757–758.
17. May M. Life Science Technologies: big biological impacts from Big Data. Science 2014; 344: 1298–1300.
18. Wang W, Krishnan E. Big Data and clinicians: a review on the state of the science. JMIR Med Informatics 2014; 2: e1.
19. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet 2012; 13: 395–405.
20. Currie J. 'Big Data' versus 'big brother': on the appropriate use of large-scale data collections in pediatrics. Pediatrics 2013; 131: S127–S132.
21. Marx V. Biology: the big challenges of Big Data. Nature 2013; 498: 255–60.
22. Apple's ResearchKit frees medical research. Nat Biotechnol 2015; 33: 322.
23. Costa FF. Social networks, web-based tools and diseases: implications for biomedical research. Drug Discov Today 2013; 18: 272–81.
24. Pereira S, Gibbs RA, McGuire AL. Open access data sharing in genomic research. Genes (Basel) 2014; 5: 739–47.
25. Weber GM, Mandl KD, Kohane IS. Finding the missing link for big biomedical data. JAMA 2014; 311: 2479–80.
26. Allen NE, Sudlow C, Peakman T, Collins R. UK biobank data: come and get it. Sci Transl Med 2014; 6: 224ed4.
27. Van Ommen G-JB, Törnwall O, Bréchot C et al. BBMRI-ERIC as a resource for pharmaceutical and life science industries: the development of biobank-based Expert Centres. Eur J Hum Genet 2015; 23: 893–900.
28. Knoppers BM. International ethics harmonization and the global alliance for genomics and health. Genome Med 2014; 6: 13.
29. Gellert R, Gutwirth S. The legal construction of privacy and data protection. Comput Law Secur Rev 2013; 29: 522–30.
30. Hallinan D, Friedewald M, De Hert P. Genetic data and the data protection regulation: anonymity, multiple subjects, sensitivity and a prohibitionary logic regarding genetic data? Comput Law Secur Rev 2013; 29: 317–29.
31. New challenges to data protection, Working Paper No. 2: Data protection laws in the EU. European Commission, 2010. Available at http://ec.europa.eu/justice/policies/privacy/docs/studies/new_privacy_challenges/final_report_working_paper_2_en.pdf.
32. Nuffield Council on Bioethics. The collection, linking and use of data in biomedical research and health care: ethical issues, 2015. Available at http://nuffieldbioethics.org/wp-content/uploads/Biological_and_health_data_web.pdf.
33. Mittelstadt BD, Floridi L. The ethics of Big Data: current and foreseeable issues in biomedical contexts. Sci Eng Ethics 2015, e-pub ahead of print 23 May 2015 doi:10.1007/s11948-015-9652-2.

34. Boddington P, Curren L, Kaye J et al. Consent forms in genomics: the difference between law and practice. Eur J Health Law 2011; 18: 491–519.
35. McGuire AL, Beskow LM. Informed consent in genomics and genetic research. Annu Rev Genomics Hum Genet 2010; 11: 361–81.
36. Steinsbekk KS, Kåre Myskja B, Solberg B. Broad consent versus dynamic consent in biobank research: is passive participation an ethical problem? Eur J Hum Genet 2013; 21: 897–902.
37. Casali PG. Risks of the new EU Data protection regulation: an ESMO position paper endorsed by the European oncology community. Ann Oncol 2014; 25: 1458–61.
38. Petrini C. 'Broad' consent, exceptions to consent and the question of using biological samples for research purposes different from the initial collection purpose. Soc Sci Med 2010; 70: 217–20.
39. Laurie G, Postan E. Rhetoric or reality: what is the legal status of the consent form in health-related research? Med Law Rev 2013; 21: 371–414.
40. Master Z, Nelson E, Murdoch B, Caulfield T. Biobanks, consent and claims of consensus. Nat Methods 2012; 9: 885–8.
41. Kaye J. The tension between data sharing and the protection of privacy in genomics research. Annu Rev Genomics Hum Genet 2012; 13: 415–31.
42. Allen J, McNamara B. Reconsidering the value of consent in biobank research. Bioethics 2011; 25: 155–66.
43. Caulfield T, Kaye J. Broad consent in biobanking: reflections on seemingly insurmountable dilemmas. Med Law Int 2009; 10: 85–100.
44. Helgesson G. In defense of broad consent. Camb Q Healthc Ethics 2012; 21: 40–50.
45. Hallinan D, Friedewald M. Open consent, biobanking and data protection law: can open consent be 'informed' under the forthcoming data protection regulation? Life Sci Soc Policy 2015; 11: 1.
46. Kaye J, Curren L, Anderson N et al. From patients to partners: participant-centric initiatives in biomedical research. Nat Rev Genet 2012; 13: 371–6.
47. Kaye J, Whitley EA, Lund D, Morrison M, Teare H, Melham K. Dynamic consent: a patient interface for twenty-first century research networks. Eur J Hum Genet 2014; 23: 141–6.
48. Williams H, Spencer K, Sanders C et al. Dynamic consent: a possible solution to improve patient confidence and trust in how electronic patient records are used in medical research. JMIR Med Informatics 2015; 3: e3.
49. Heeney C, Hawkins N, de Vries J, Boddington P, Kaye J. Assessing the privacy risks of data sharing in genomics. Public Health Genomics 2011; 14: 17–25.
50. Mascalzoni D, Dove ES, Rubinstein Y et al. International Charter of principles for sharing bio-specimens and data. Eur J Hum Genet 2014; 23: 721–8.
51. McGuire AL, Caulfield T, Cho MK. Research ethics and the challenge of whole-genome sequencing. Nat Rev Genet 2008; 9: 152–6.
52. Rodriguez LL, Brooks LD, Greenberg JH, Green ED. Research ethics. The complexities of genomic identifiability. Science 2013; 339: 275–6.

53. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. Science 2013; 339: 321–4.
54. Article 29 Data Protection Working Party. Opinion 05/2014 on Anonymisation Techniques, 2014. Available at http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.
55. O'Brien SJ. Stewardship of human biospecimens, DNA, genotype, and clinical data in the GWAS era. Annu Rev Genomics Hum Genet 2009; 10: 193–209.
56. Knoppers BM, Zawati MH, Kirby ES. Sampling populations of humans across the world: ELSI issues. Annu Rev Genomics Hum Genet 2012; 13: 395–413.
57. Tene O, Polonetsky J. Privacy in the age of Big Data: a time for big decisions. Stanford Law Rev Online 2012; 64: 63–9.
58. Lowrance WW. Privacy, Confidentiality, and Health Research. Cambridge University Press: Cambridge, UK, 2012.
59. van Veen EB. Europe and tissue research: a regulatory patchwork. Diagn Histopathol 2013; 19: 331–6.
60. Ploem MC. Towards an appropriate privacy regime for medical data research. Eur J Health Law 2006; 13: 41–63.
61. Sethi N. The promotion of data sharing in pharmacoepidemiology. Eur J Health Law 2014; 21: 271–96.
62. Greely HT. The uneasy ethical and legal underpinnings of large-scale genomic biobanks. Annu Rev Genomics Hum Genet 2007; 8: 343–64.
63. Laurie G. Genetic privacy: A Challenge to Medico-Legal Norms. Oxford University Press: New York, 2002.
64. Gaye A, Marcon Y, Isaeva J et al. DataSHIELD: taking the analysis to the data, not the data to the analysis. Int J Epidemiol 2014; 43: 1929–44.
65. Wolfson M, Wallace SE, Masca N et al. DataSHIELD: resolving a conflict in contemporary bioscience–performing a pooled analysis of individual-level data without sharing the data. Int J Epidemiol 2010; 39: 1372–82.
66. Wallace SE, Gaye A, Shoush O, Burton PR. Protecting personal data in epidemiological research: DataSHIELD and UK Law. Public Health Genomics 2014; 17: 149–57.
67. Budin-Ljøsne I, Burton P, Isaeva J et al. DataSHIELD: an ethically robust solution to multiple-site individual-level data analysis. Public Health Genomics 2015; 18: 87–96.
68. Abbing HD. EU cross-border healthcare and health law. Eur J Health Law 2015; 22: 1–12.
69. Prainsack B, Buyx A. A solidarity-based approach to the governance of research biobanks. Med Law Rev 2013; 21: 71–91.
70. Kaye J. From single biobanks to international networks: developing e-governance. Hum Genet 2011; 130: 377–82.

Chapter 4

A new regulatory landscape for Big Data health research: Safeguards and research exemptions in the GDPR

Mostert M.
Big data, medisch-wetenschappelijk onderzoek en gegevensbescherming.

in: Ottes L., Kits PM, Zwenne GJ, Steenbruggen WAM, van Veen EB, Kleefman TG, Miedema F, Mostert M. Big data in de zorg, preadvies uitgebracht voor de Vereniging voor Gezondheidsrecht, Sdu Uitgevers: Den Haag, 2017, p. 165-93.

Abstract

The legal framework applicable to health research using Big Data is undergoing significant changes. As from 25 May 2018, the possibilities for linking and analysing personal data in research will largely be determined by the General Data Protection Regulation (GDPR), and by its interpretation and implementation in national law. In this Chapter, I examine whether the EU legislator has achieved its dual objective of both facilitating health research and subjecting it to appropriate safeguards. Key issues in this analysis are anonymity, the informed consent requirement, and research exemptions related to informed consent, key data processing principles and individual rights. Although the EU legislator has not achieved a high level of harmonisation, the GDPR does offer sufficient leeway for a specific regime for health research. Further attention should be paid to the elaboration on appropriate governance measures, which could be aligned with recent internationally recognised ethical guidelines.

Introduction

Across the globe, investments are steadily increasing to improve knowledge about human health and disease with the help of Big Data.[1,2] The current use of Big Data in health research primarily consists of collecting, reusing, linking and analysing diverse, large, and/or complex data files. This development is also called data-intensive health research.[3] The analyses are carried out using both traditional and innovative research methods. As of 25 May 2018, the use of personal data in such research in the European Union (EU) is subject to the provisions set out in the General Data Protection Regulation (GDPR) (EU 2016/679). Given the comprehensive protection provided for in the GDPR and the fact that this regulation will be directly applicable in all EU Member States, this is a very important development. Despite the GDPR's direct effect, a substantial part of its provisions relating to scientific research do need to be implemented in national law. The implementation of the GDPR in the Netherlands will occur through the Dutch GDPR Implementation Act ('Uitvoeringswet Algemene Verordening Gegevensbescherming', UAVG), the consultation version of which was recently published.[4] When this GDPR Implementation Act enters into force, the Dutch Personal Data Protection Act ('Wet bescherming persoonsgegevens', Wbp) will be repealed.

The importance of data-intensive health research, or at least data linkage, is explicitly recognised in the GDPR. The EU legislator considers that 'researchers can obtain new knowledge of great value with regard to widespread medical conditions such as cardiovascular disease, cancer and depression' by linking data. Therefore, and still according to the EU legislator, scientific research in the EU should be facilitated by determining that 'In order to facilitate scientific research, personal data can be processed for scientific research purposes, subject to appropriate conditions and safeguards (..)' (Recital 157 GDPR). Reference is also made to the EU's objective of achieving a 'European Research Area' as set out in Article 179(1) of the Treaty on the Functioning of the European Union (Recital 159 GDPR). Despite these intentions, after the European Parliament voted in favour of a draft version of the GDPR, it was feared that the GDPR would seriously impede health research.[5] The amended final version of the GDPR, however, received a more positive response. A message was soon published in The Lancet stating that the GDPR represents a step forward, both in the protection of personal data and in its use in biomedical research.[6]

However, shortly afterwards, a brief article-by-article analysis commissioned by The Wellcome Trust showed that the GDPR's impact on research depends to a

large extent on the interpretation and implementation of the GDPR in the various Member States.[7] There is often room for interpretation when it comes to the definition of personal data, the conditions for legally valid consent, the exceptions to the consent requirement, the possible derogations from various principles and rights, and the appropriate safeguards required in the event of exceptions or derogations. Therefore, this contribution aims to follow up on this brief analysis and provide a more detailed analysis and deliberation. The central question is whether the GDPR contributes positively to the responsible and effective use of personal data in data-intensive health research. Furthermore, we will also take into consideration the consultation version of the Dutch GDPR Implementation Act.

From privacy to data protection

Although the EU legislator wishes to facilitate the processing of personal data for research purposes, it is stated first and foremost that this should be subject to appropriate conditions and safeguards (Recital 157 GDPR). Essentially, such safeguards serve to ensure adequate protection of the fundamental rights of natural persons. Yet interestingly, it is precisely at the level of those fundamental rights that a change is occurring which is reflected in the GDPR. While Directive 95/46/EC specifically guaranteed the right to privacy (see Article 1(1) of the Directive), the GDPR is focused largely on guaranteeing the right to data protection. This is apparent from provisions such as Article 1(2) GDPR, which states that the regulation aims to protect the fundamental rights of natural persons 'and in particular their right to the protection of personal data'. Equally striking is that the use of the term 'privacy' has all but disappeared from (the English version of) the GDPR. Common terms such as 'privacy by design' and 'privacy impact assessment' have been changed to 'data protection by design' and 'data protection impact assessment'. Again, the EU legislator appears to convey that the protection of personal data no longer primarily guarantees the right to privacy.

These amendments to the GDPR follow the recognition of the right to data protection as a fundamental right in the EU, distinct from and complementary to the right to privacy. The recognition of the right to data protection is enshrined in Article 8 of the EU Charter of Fundamental Rights (EU Charter), while the recognition of the right to privacy is set out in Article 7 of the EU Charter. Since 2009, the EU Charter has been legally binding for the EU and its Member States in situations where they when they are implementing EU law.[8] The EU Charter is therefore of

particular relevance to the interpretation and implementation of the GDPR, taking precedence over national law and therefore over Article 10 of the Dutch Constitution. In its limited scope, the EU Charter differs from the European Convention on Human Rights (ECHR), which has a more extensive application in the national legal order. Despite this limitation, the right to data protection set out in Art. 8 of the EU Charter partly provides a higher level of protection than the right to privacy as recognised in Article 8 of the ECHR and Article 7 of the EU Charter. According to the Dutch Interdepartmental Committee on European Law (ICER), this extension of the level of protection consists in the recognition of the right to independent supervision in Article 8(3) of the EU Charter.[8] It should also be noted in this regard that ensuring the right to data protection is essentially a positive obligation for the EU and its Member States to subject the processing of personal data to appropriate rules. This applies both to data processing in horizontal relationships and to data processing by public authorities in relation to citizens. The right to privacy is different in nature. Observing the right to privacy still remains, at its core, a negative obligation for public authorities.[9] We should note, however, that Article 8 of the ECHR increasingly entails a positive obligation to guarantee the right to privacy, including between citizens.[10] Nevertheless, based on the fundamental right to data protection, the EU and its Member States can be expected to fulfil a more active role in protecting the flow of personal data in accordance with the requirements set out in Article 8 of the EU Charter. The motive for this protection also includes the promotion of collective interests,[11] such as responsibly enabling data exchange for scientific research purposes.

The EU legislator stresses that the interest in protecting personal data is increasing in light of rapid technological and societal changes (Recital 6 GDPR). Such changes are definitely also occurring when it comes to the processing of health data and other special categories of personal data. The days when it was mostly healthcare providers who had access to health data and recorded them in the medical file are now long gone. Increasingly, data about an individual's health, genes and lifestyle are also recorded, shared and/or reused on a large scale by other parties than the healthcare provider, thereby no longer falling within the scope of medical confidentiality. Examples of developments contributing to this, specifically focused on scientific research, include the rise of *citizen science* and *participant-driven research*,[12] but certainly also applications such as Apple's ResearchKit, used increasingly by scientists.[13] In addition, the patient's right to receive a digital copy of their medical file on the basis of a new Dutch law may also contribute to the dissemination of

patient data unbeknownst to the healthcare provider.[14] This further increases the importance of an overarching data protection regime to complement laws related to medical confidentiality.[15]

Identifiability in the Big Data age

The lawful use of personal data for health research purposes is regularly made possible in practice by anonymising the data. It can also happen that personal data are wrongfully regarded as anonymous.[16] This is an attractive avenue, as the legal regime for data protection only applies to the processing of personal data. 'Personal data' refers to any information that can be used to identify a natural person, either directly or indirectly (Article 4(1) of the GDPR).[17] Article 89(1) of the GDPR even prescribes the use of anonymous data if the research objectives can be achieved in this way. Although at first sight anonymisation appears to be a useful strategy for scientists to meet legal requirements, critical reflection on the effectiveness of anonymisation techniques is in order. And it is all the more apropos now that, with the advent of Big Data, the possibilities for profiling and re-identification are increasing.

The end of anonymity?

It is clear that absolute anonymity can no longer be guaranteed,[18] especially when it comes to genetic data.[19–21] However, absolute anonymity is not required by law. According to Recital 26 of the GDPR, a test based in reason determines whether data are anonymous: 'To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly'. Data are therefore considered anonymous whenever direct or indirect identification is prevented, based on the means that persons can *reasonably likely* deploy.

In implementing this norm, as per the aforementioned GDPR Recital, one must not only take into account current technology, but also 'technological developments'. For the time being, these technological developments mostly seem to be contributing to an increased risk of re-identification. Opinion 5/2014 of the Article 29 Working Party, for instance, showed that none of the anonymisation techniques analysed met the criteria for effective anonymisation with certainty.[22] According to the Article 29 Working Party, this is not solely to do with technical factors; relevant contextual factors should also be taken into account, such as measures restricting access to

data or the availability of public sources of information. We should also factor in that the amount of information that is publicly available, or that might otherwise become available with reasonable effort, is growing exponentially. In addition, the possibilities for analysis are increasing, so that a combination of available sources may lead to prompt identification.[23]

A recent judgment by the Court of Justice of the European Union (CJEU) clarified when externally available information should be considered as a reasonably deployable means of identification.[24] According to the Court, this is not the case if identification by means of this additional information is 'practically impossible on account of the fact that it requires a disproportionate effort in terms of time, cost and man-power, so that the risk of identification appears in reality to be insignificant'. In addition, the Court considers that a statutory prohibition – rendering access to the extra information legally impossible – might also entail that said extra information is no longer to be regarded as a reasonably likely deployable means to identify a person.

Although compliance with the aforementioned legal standard of anonymity is not impossible, effective anonymisation does significantly reduce the usefulness of the data for research purposes. Firstly, this is due to the fact that a lot of useful data has to be removed from a file in order to reasonably avoid identification of the individual. Secondly, anonymisation seems to render impossible any further linking or updating of these data files. This would seriously hamper data-intensive health research, as the possibility of linking and updating data is one of the essential elements of such research.[3,25–27]

Risk reduction through pseudonymisation

Although most of the disadvantages of the use of anonymous data also apply to the use of pseudonymous data (as defined in Article 4(5) GDPR), certain possibilities for linking and updating the pseudonymised data do remain.[12] Partly for this reason, some proposed that pseudonymous (double-coded) data used for research purposes should be regarded as anonymous.[28,29] However, the GDPR stresses that, while pseudonymisation is a risk reduction tool, it is not intended to exclude other data protection measures (Recital 28 GDPR). It would seem that in doing so, the EU legislator unambiguously confirms that pseudonymous data should be considered as personal data.

Nevertheless, the aforementioned CJEU ruling leaves some room for doubt, if and to the extent that a ban on indirect identification were to be introduced by law. This would need to be a ban that renders access to the extra information required

to trace the pseudonymous data back to the individual impossible from a legal point of view.[24]

The status of human samples

Genetic data are protected by the GDPR as long as they meet the definition in Article 4(13) of the GDPR. From this paragraph it can be deduced that the human sample itself does not fall within the definition of genetic data. After all, the definition mentions data 'which result, in particular, from an analysis of a biological sample from the natural person in question.' This is confirmed in Recital 34 of the GDPR. However, the Explanatory Memorandum to the consultation version of the Dutch GDPR Implementation Act contains the following divergent definition: 'This is understood to mean, among other things, the DNA of a person or material from which information relating to the DNA can be derived'. According to this definition, the body material itself, from which information relating to the DNA can be derived, also falls within the definition of genetic data.

Anyhow, the storage of human samples should *in itself* be considered as an infringement of *the right to privacy*. This is shown by the following Recital of the European Court of Human Rights (ECtHR) in the case of S. and Marper v the United Kingdom: 'Given the nature and the amount of personal information contained in cellular samples, their retention per se must be regarded as interfering with the right to respect for the private lives of the individuals concerned.'[30] Furthermore, the ECtHR also concludes that the storage of human samples in particular infringes the right to privacy, 'given the wealth of genetic and health information contained therein.'[31] Ploem agrees with this conclusion, because she believes that the (future) potential to derive sensitive information from it is greater for human samples than for a combination of data files.[15]

Subconclusion

We have now shown that the use of anonymisation techniques is often not a sufficiently reliable strategy to protect the interests and rights of data subjects. Even after data have been declared anonymous, there is still a need to monitor this anonymity and to control the context in which the data exist. In addition, truly anonymous databases are virtually unusable for data-intensive health research purposes. Any claim that data files that can be used for health research are anonymous should therefore be examined critically, especially when it comes to genetic data or publicly available data files.

Although the storage of human samples should be considered as an infringement of the right to privacy, the actual sample itself does not seem to fall within the GDPR's definition of personal data or genetic data. In my opinion, the differing definition in the consultation version of the Dutch GDPR Implementation Act should therefore be amended or further clarified. Should the Dutch legislator decide to maintain the aforementioned definition, then it would at least be important to make a distinction between human samples that should be considered as personal data and samples that should not be. This distinction could be made on the basis of the purpose for which the human samples are being stored. Whenever the purpose of the storage is to derive information from the samples and it is reasonably possible to trace this information back to the individual, one might consider classifying the human samples as personal data. Given the increasing potential for deriving special personal data from human samples, this would be beneficial to the protection of the individual. However, on the basis of the definitions in the GDPR, I conclude that personal data are not processed until such time as they are derived from the human samples by means of an analysis. Thus, the need for specific legislation governing the use of human samples remains, also in light of the special nature and sensitive status of human samples.

Consent: specific or broad?

In the event that anonymisation does not appear to be a desirable strategy for using data in data-intensive health research, asking the data subjects for their consent is a possible avenue, allowing for personal data to be processed lawfully in data-intensive health research. A widely-held view in literature, however, is that strict adherence to the consent requirement hinders – or even renders impossible – the effective use of personal data in data-intensive health research.[3] The situation may prove difficult or unworkable, especially if the consent given is limited to very specific purposes, meaning that in the event of reuse or linking, consent must be obtained again and again. The elaboration and interpretation of the consent requirement, and in particular the extent to which the purposes of the processing should be specified, is therefore of great importance.

Disentangling consent requirements

In general, data-intensive health research will have to comply with the specific regime in Article 9 of the GDPR, because it can be assumed that such research can

rarely do without the processing of any special (categories of) personal data, such as health data. Special categories of personal data may not be processed except in the cases expressly referred to in Article 9 of the GDPR, requiring the explicit consent of the data subject in accordance with Article 9(2)(a) of the GDPR. The prohibition and the specific requirements in Article 9 of the GDPR are complementary to the general principles and rules of the GDPR. Exceptions to the prohibition in Article 9 of the GDPR are only permitted under the specific conditions that are expressly mentioned (Recital 51 GDPR). The requirement for explicit consent in Article 9(2) of the GDPR is a special provision that applies in addition to, and takes precedence over, the general consent requirement. It is therefore important to distinguish between these two different consent requirements.

The general consent requirement is defined in Article 4(11) of the GDPR as a 'freely given, specific, informed and unambiguous indication' by which the data subject agrees to the processing. Pursuant to Article 6(1)(a) of the GDPR, consent must be given for one or more specific purposes. As explained above, adhering in full to the latter requirement of purpose limitation would constitute a major impediment to data-intensive health research. The EU legislator therefore considers that it is often not possible to comprehensively define the purpose of the data processing at the moment of collection of the personal data. The following is then put forward: 'Therefore, 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.' (Recital 33 GDPR). The consent for certain research areas as described here is quite comparable with so-called 'broad consent'.

Hence, Recital 33 of the GDPR provides a starting point for national legislators and policymakers to regard broad consent as a lawful form of consent within the meaning of Article 6(1)(a) of the GDPR. However, the EU legislator explicitly considers that established ethical standards for scientific research should then be respected. A recent document setting out such ethical standards is the World Medical Association Declaration of Taipei on Ethical Considerations regarding Health Databases and Biobanks (WMA Declaration of Taipei).[32] According to the WMA Declaration of Taipei, broad consent is only valid if the individual has been informed about the issues mentioned in paragraph 12, including the governance measures mentioned in paragraph 21. It should be noted, however, that the Declaration of Taipei is primarily aimed at physicians. The also recently established Council for International Organizations of Medical Sciences International Ethical Guidelines for Health-related Research Involving Humans (CIOMS Guidelines) are aimed at a broader

audience.[33] Guideline 12 of the CIOMS Guidelines confirms that broad consent is an acceptable alternative, as long as it is accompanied by appropriate governance of the data collection. It is emphasised that broad consent is essentially consent based on (information about) an adequate governance system.[34]

Whenever special categories of personal data are processed for research purposes, the provisions of Article 9(2)(a) of the GDPR apply, in addition to or in deviation from the general consent requirement. Said subsection calls for 'explicit' consent, given for 'one or more specified purposes. The explicit consent pursuant to Article 9(2)(a) of the GDPR must therefore relate to the processing for specified or precisely defined purposes.

The question is then whether broad consent (for specific research purposes) also meets the consent requirement set out in Article 9(2)(a) of the GDPR. Neither the specific rules stipulated in Article 9 of the GDPR, nor the Recitals related to this article provide for any derogation from the requirement of specificity or purpose limitation. In addition, Recital 33 of the GDPR does not explicitly address the processing of special categories of personal data; there is no reference whatsoever to Article 9 of the GDPR or to the specific regime for such data. It should be borne in mind that deviations from the prohibition in Article 9 of the GDPR are only permitted insofar as they are explicitly provided for. However, the consultation version of the GDPR Implementation Act shows that the Dutch Ministry of Security and Justice assumes that Recital 33 of the GDPR does also apply to the requirement for explicit consent. This is unequivocally confirmed in the Explanatory Memorandum to the consultation version.[4]

Subconclusion

Although the EU legislator allows for broad consent with regard to the processing of personal data for research purposes, it is unclear whether this broadening also fully applies to the processing of special categories of personal data. Article 9 of the GDPR provides for a specific regime that, in itself, does not explicitly allow for a derogation from the specific consent requirement. If this were to mean that broad consent is not permitted for the processing of special categories of personal data, this would considerably limit the possibilities of obtaining legally valid consent in data-intensive health research. It is unclear whether this is the EU legislator's intention. The interpretation in the consultation version of the GDPR Implementation Act is therefore also a justifiable one. In the interests of legal certainty, however, it is important for the European Commission or another authority to clarify this issue.

If it can be assumed that broad consent is sufficient for the processing of special categories of personal data for research purposes, then the processing must at least meet the requirements set out in Recital 33 of the GDPR. This means that broad consent must be focused on certain areas of scientific research and that established ethical standards must be observed. The lawfulness of broad consent therefore depends in part on compliance with such established ethical standards. It follows that established ethical standards, as laid down in the CIOMS Guidelines and the WMA Declaration of Taipei, are given more legal significance.

Alternative to the consent requirement

As an alternative to the consent requirement, Member State or EU law may contain a provision that specifically exempts the processing of special categories of personal data for research purposes from the prohibition in Article 9 of the GDPR (Article 9 (2)(j) GDPR). Although the importance of such an exemption is hardly disputed, there is some contention as to its proper conditions from a legal and ethical point of view.[3] Against this background, what follows is a reflection on the conditions attached by Article 9(2)(j) of the GDPR to any exemption from the consent requirement implemented by national law. We will also discuss how this provision is to be implemented into Dutch law according to the consultation version of the GDPR Implementation Act.

Necessity and proportionality

Firstly, pursuant to Article 9(2)(j) of the GDPR, data processing must be both necessary and proportionate to the research purposes. This indicates the importance of the principles of necessity and proportionality in the GDPR. The European Parliament and the Council of Ministers have taken different views on the proper interpretation of these fundamental principles. On 12 March 2014, the European Parliament voted in favour of a draft version of the GDPR that only allowed exceptions for research serving a '*high public interest*'.[35] It was feared that this requirement would seriously impede health research.[5] Later, the Council of Ministers agreed on a draft version of the GDPR that allowed significantly more room for exception provisions.[36] Ultimately, a compromise was reached, resulting in the above-mentioned open standards, which require further elaboration.

Many different views have been recorded in literature on the proper further elaboration of an exemption from the consent requirement for medical-scientific research.[3] There are also significant differences between the ethical standards

recognised in the CIOMS Guidelines and those in the WMA Declaration of Taipei. According to paragraph 16 of the WMA Declaration of Taipei, an exception to the consent requirement is only allowed with the prior approval of an Ethics Committee. In addition, this Ethics Committee can only override the consent requirement in the event of a concrete, serious and immediate threat to public health. In imposing this condition, the WMA appears to convey the opinion that the importance of advancing medical science, in itself, is never sufficient to override the consent requirement.

The CIOMS Guidelines are considerably less strict, referring to two situations where exceptions to the consent requirement are possible according to Guideline 12. The first situation concerns the use of data that are already being collected in the context of routine clinical care. In such cases, an informed opt-out procedure is sufficient according to CIOMS. This opt-out procedure must meet the following conditions: (a) patients must be aware of the procedure; (b) patients must be provided with sufficient information; (c) patients must be made aware of the right to withdraw their data; and (d) patients must have a real possibility to object. It should be noted here that, in the context of clinical care, such data are often governed by medical confidentiality, which is subject to a stricter legal regime. In the Netherlands, it is difficult to reconcile the principle of an opt-out procedure with the current regime. After all, pursuant to Article 7:458 of the Dutch Civil Code, consent is to be obtained unless this is not reasonably possible or cannot reasonably be required. A no-objection system similar to the aforementioned opt-out procedure applies only after one of these conditions has been met. The second situation described in Guideline 12 of the CIOMS Guidelines is when researchers wish to use data collected in the past for research purposes, without consent having been given for future use. In such a case, an Ethics Committee may grant derogations from the consent requirement, if all of the following conditions are met. The study must: (a) not be feasible or practicable if consent is to be obtained; (b) have 'important social value'; and (c) not pose more than a minimal risk to the participants or the group to which they belong.

The interpretation of the aforementioned principles in the consultation version of the GDPR Implementation Act is similar in nature to that of Article 23(2) of the current Dutch Data Protection Act (Wbp). According to Article 27 of the consultation version of the GDPR Implementation Act, the processing must be necessary for research that serves a public interest. Given that scientific research often serves public interest, this requirement is not likely to entail any significant restrictions. However, the principle of subsidiarity linked to the necessity requirement compels us to verify whether the research objective cannot be achieved in another way that is less

intrusive for the data subjects. Therefore, as stated in Article 27 of the consultation version of the GDPR Implementation Act, the sole case where this prohibition does not apply is when obtaining explicit consent either proves to be impossible or requires a disproportionate effort. According to the Explanatory Memorandum to the Dutch Data Protection Act (Wbp), whether or not a disproportionate effort would be required depends in part on the extent to which possible avenues to inform the data subjects are available. Consideration should be given to whether or not a medium is available through which a large majority of the data subjects can be reached.[37] Such mediums include portals or interfaces that enable digital communication with large groups of (potential) participants, an example of which is dynamic consent.[38]

Respecting the right to data protection

Secondly, and still pursuant to Article 9(2)(j) of the GDPR, 'the essence of the right to data protection' must be respected. A concrete explanation of how this condition is (to be) met appears to be lacking in the consultation version of the GDPR Implementation Act. This does not contribute to legal certainty in this regard.

While it is generally not a simple matter to determine the essential content or essence of a fundamental right, the wording of Article 8 of the EU Charter contains fairly concrete indications. In accordance with Article 8(2) of the EU Charter, the essential content includes the principles of fairness, purpose limitation and lawfulness. The rights of access to and rectification of personal data must also be observed. In addition, Article 8(3) of the EU Charter calls for independent supervision. Furthermore, other elements may also be considered as essential content of the right to data protection on the basis of Article 8(1) of the EU Charter.

The specific rules for the processing of special categories of personal data therefore seem to require, among other things, that the purpose limitation principle be observed if an exception is granted on the grounds of Article 9(2)(j) of the GDPR. Indeed, the purpose limitation principle is part of the essential content of the right to data protection. However, it is unclear how this requirement, which stems from the specific regime of protection under Article 9 of the GDPR, relates to the general rules governing the processing of personal data. The general rules provide that, under certain conditions, further processing for research purposes is not subject to the purpose limitation principle (Article 5(1)(b) of the GDPR). Nevertheless, this exception to the purpose limitation principle does not explicitly address the processing of special categories of personal data. Neither in Article 5 of the GDPR, nor in any of the Recitals related to it is there any reference to the specific regime

of protection in accordance with Article 9 of the GDPR. As explicit provision must be made for derogations from the prohibition in Article 9 of the GDPR, it is unclear whether the purpose limitation principle is to be (fully) observed.

Safeguards under Article 89(1) of the GDPR

Thirdly, the exception provision must be subject to the appropriate safeguards as set out in Article 89(1) of the GDPR. According to the first sentence of this paragraph, the safeguards must be in accordance with the GDPR and the rights and freedoms of the data subject. In the Dutch version of the GDPR, the wording seems to indicate that the safeguards required under Article 89(1) of the GDPR are limited to governance measures ensuring the principle of data minimisation. These measures 'may' include pseudonymisation or anonymisation, as long as this does not impede the achievement of the research objectives. This would therefore appear to be an optional rule from which it would be easy to derogate in the interests of research. In addition, as per Article 5(1)(c) of the GDPR, the processing must already adhere to the principle of data minimisation anyway. In its Dutch translation, Article 89(1) of the GDPR seems to add little of relevance to the data protection regime that is already in place. In the English version of the GDPR, by contrast, the second sentence of Article 89(1) reads as follows: "Those safeguards shall ensure that technical and organisational measures are in place in particular *[underlined by author]* in order to ensure respect for the principle of data minimisation." In other words, while these governance measures need to ensure the principle of data minimisation, they are not necessarily limited to this. Unfortunately, neither Art. 89(1) of the GDPR, nor the Recitals related to it contain any concrete indications about other possible governance measures.

From an ethical perspective, the necessary governance measures have become clearer. The CIOMS Guidelines and the WMA Declaration of Taipei contain relatively concrete and specific standards aimed at establishing an appropriate governance structure at the institutional level. Although these standards are not legally binding, they are intended to contribute to legislation and regulations in this domain. The CIOMS Guidelines and the WMA Declaration of Taipei complement the legal standards in the GDPR by stating, among other things, that the governance structure must include: mechanisms to get back in touch with the individual; the revelation of random findings and the communication of scientific findings; a body that assesses data releases for research and the conditions under which they occur; and participation by patient groups or the wider community.[32,33]

Art. 89(1) of the GDPR directly affects the Dutch legal order and does not require implementation into national law. The conditions of Article 89(1) of the GDPR must therefore be met, in addition to those in Article 27 of the consultation version of the GDPR Implementation Act. It is advisable to mention this explicitly in the explanatory notes to the GDPR Implementation Act.

Additional specific measures

Fourthly, appropriate and specific measures must be taken to protect the fundamental rights and interests of the data subject, in addition to those already required under Article 89(1) of the GDPR. Article 9(2)(j) of the GDPR emphasises that measures based on Article 89(1) of the GDPR alone are not sufficient. Appropriate measures should be taken, specifically aimed at an alternative to the consent requirement. However, the GDPR does not provide any indication as to what these appropriate and specific measures should be, or how their implementation should be ensured. These aspects may be further specified later in EU or national law.

The consultation version of the GDPR Implementation Act does not specify which additional specific measures should be taken. This condition seems to be implicitly included in the general requirement in Article 27(c) of the consultation version.

Processing of genetic data, additional conditions

Article 9(4) of the GDPR allows Member States to impose additional conditions for the processing of genetic (and other) data. In the consultation version of the GDPR Implementation Act, such additional conditions are set forth in Article 24. According to this Article, genetic data may be processed only in two cases, namely: (a) if substantial medical interests so require; or (b) if the processing is necessary for scientific research or statistical purposes *and* the data subject has given their explicit consent. Based on this provision, the use of genetic data in scientific research would only be allowed after having obtained explicit consent. Although this is a topic of debate,[39] this relatively strict regime for the processing of genetic data is based on the idea that genetic data need extra protection compared to other special categories of personal data.

Subconclusion

In light of the above, it has become apparent that Article 9(2)(j) of the GDPR mainly contains open standards that require further elaboration in EU or Member State law. Firstly, the principles of necessity and proportionality must be observed. The

legislative process, the literature and established ethical standards show strongly diverging views on how these standards should be interpreted. It is therefore likely that a different regime will be established in each Member State, unless EU law provides for a supplementary regime after all. On the basis of the consultation version of the GDPR Implementation Act, asking for consent would remain the preferred option in the Netherlands, as exceptions are only permitted when obtaining explicit consent either proves to be impossible or requires a disproportionate effort. In interpreting these standards, consideration should be given to digital media that could be used to reach large groups of people, such as communication through e-mail, portals or other interfaces. A large study population is therefore not, in and of itself, a sufficient reason to assume that obtaining consent would require a disproportionate effort.

Secondly, any exception provision must be subject to appropriate safeguards and specific measures. The only concrete measures or safeguards mentioned in the GDPR are listed in Article 89(1) of the GDPR: pseudonymisation or anonymisation of data wherever possible. In addition, specific measures not specified in the GDPR must be taken, and the essential content of the right to data protection must be respected. The consultation version of the GDPR Implementation Act does not contain any concrete mention of which necessary additional appropriate and specific measures should be taken either. This lack of harmonisation will not facilitate the exchange of special categories of personal data between EU Member States without the consent of the data subject. On the other hand, the open standards do offer some leeway to develop flexible standards at the national level, allowing not only legislation, but also (ethical) guidelines to play an important role. When implementing appropriate and specific measures, one could envisage maintaining an easily accessible and effective no-objection system, thereby facilitating the exercise of the right of objection on the grounds of Article 21 of the GDPR. Another option to consider is prior review by an Ethics Committee, as recommended by the CIOMS Guidelines and the Declaration of Taipei. This may be particularly useful when it comes to providing access to data sets that, given their nature or size, deserve special protection.

Principles and individual Rights

While the above considerations were mainly focused on the principle of lawfulness, there are other principles, such as purpose limitation and storage limitation, that could conflict with the effective use of Big Data in health research. The same goes for a number of rights of data subjects under the general data protection regime. In

order to prevent processing for research purposes from being restricted by this to a disproportionate extent, the GDPR leaves some room for derogations from several of these principles and rights.

Derogations from principles for research

The reuse of personal data in health research is facilitated by derogations from the principles of purpose limitation and storage limitation. On the grounds of Article 5(1)(b) of the GDPR, further processing for research purposes is not regarded as incompatible with the original purposes. Furthermore, Article 5(1)(e) of the GDPR states that data used for research purposes may be stored for longer periods of time. For both of these exceptions, the processing must be in accordance with Article 89(1) of the GDPR, and the required technical and organisational measures must be taken.

Processing for research purposes is not explicitly excluded from the other principles in Article 5 of the GDPR. This means that the principles of lawfulness, fairness and transparency, data minimisation, accuracy, integrity, confidentiality and accountability must be upheld when processing data for research purposes. Nevertheless, there are various possible derogations from individual rights of data subjects based on the above principles.

Rights with derogations for research

Pursuant to Article 89(2) of the GDPR, some individual rights may be deviated from in the interests of research, if EU law or national law provides a basis for doing so. Specifically, these rights are the right of access (Article 15 GDPR), the right to rectification (Article 16 GDPR), the right to restriction of processing (Article 18 GDPR) and the right to object (Article 21 GDPR). In addition, the GDPR itself provides the basis for derogations from two other individual rights: 'the right to be forgotten' or the right to erasure (Article 17 (3)(d) GDPR) and the right to information as laid down in Article 14 GDPR (Article 14 (5)(b) GDPR). The latter right applies when the data have not been obtained directly from the data subject.

All such derogations from individual rights must comply with the principles of necessity and proportionality. Derogations are only allowed if and insofar as the individual rights would render impossible or seriously impede the processing for research purposes (Articles 14(5)(b), 17(3)(d) and 89(2) GDPR). Furthermore, any deviation must be in accordance with Article 89(1) of the GDPR, and the measures set out in this paragraph must be taken. Specifically with regard to the derogation from the right to information as per Article 14 of the GDPR, additional appropriate

measures must be taken, 'including making the information publicly available'. Such disclosure of information can only be done in a general manner, for example on a website.

In the consultation version of the GDPR Implementation Act, the content of Article 44 of the Dutch Data Protection Act (Wbp) has been maintained, save for a linguistic change. As a result, derogations from Articles 15, 16 and 18 of the GDPR are allowed.[4] Although the conditions set out in Article 89(1) and (2) of the GDPR must also be met, this is not explicitly emphasised in the consultation version of the GDPR Implementation Act. Quite deliberately, no mention is made of any derogation from the right of objection in Article 21 of the GDPR. This would mean that data subjects could fully object, under Article 21 of the GDPR, on grounds relating to their specific situation. Pursuant to Article 21(6) of the GDPR, derogations from this right for research purposes are only allowed if necessary for the performance of a task carried out in the public interest.

Rights without derogations for research

There are several individual rights with no specific derogations allowed for processing for research purposes: the right to information pursuant to Article 13 of the GDPR, the obligation to notify (Article 19 GDPR), the right to data portability (Article 20 GDPR) and the right not to be subject to a decision based solely on automated processing (Article 22 GDPR).

Particular attention should be paid to the right to information as laid down in Article 13 of the GDPR. This provision establishes that the controller has the duty to provide information in cases where the data are collected from the data subjects themselves. This includes information about the controller and the data protection officer, the purposes of and the legal basis for the processing, the recipients of the personal data, the storage period, and the rights of the data subject (Article 13(1)(2) GDPR). According to Article 13 of the GDPR, the only case in which this obligation to provide information does not apply is where the data subject already possesses the aforementioned information. However, it is somewhat confusing that, at first sight, Recital 62 of the GDPR seems to allow for a wider exemption for research purposes. This Recital states that the provision of information is not necessary if this proves to be impossible or would involve a disproportionate effort – in particular as regards processing for research purposes – yet it is unclear whether this Recital also relates to the obligation to provide information under Article 13 of the GDPR. In my opinion, interpretations in accordance with this Recital are not supported

by the text of Article 13 of the GDPR. Instead, this Recital seems rather to relate to the derogation from the obligation to provide information under Article 14 of the GDPR, as set out in Article 14(5)(b) of the GDPR.

In principle, whenever data are further processed for another purpose, all relevant information must be provided to the data subject in advance, as per Article 13(3) of the GDPR. In the case of further processing for research purposes, however, it is justifiable that this obligation to provide information does not apply in full. After all, the derogation from the principle of purpose limitation in Article 5(1)(b) of the GDPR is based on the fiction that further processing for scientific research is compatible with the original purposes. Reuse for (other) research purposes therefore does not seem to qualify as further processing for another purpose as referred to in Article 13(3) of the GDPR. In this case, the requirement applies that the further processing must comply with the conditions in Article 89(1) of the GDPR. Furthermore, it is unclear whether the obligation to provide information continues after the data have been collected, in the event of changes to the aforementioned details, such as the controller's contact details. If there is a continuous obligation to provide detailed information, this can place a significant administrative burden on organisations that retain data for long periods of time.[7]

Subconclusion

A discussion solely limited to the subjects of anonymity and (exemptions from) the consent requirement would not do justice to the more comprehensive regime of data protection. After all, the principle of lawfulness is only one element of this regime, in addition to the other principles referred to in Article 5 of the GDPR and the rules and individual rights based on those principles. The derogations from some of these general principles and individual rights seem to sufficiently facilitate the use of personal data in data-intensive health research. Nevertheless, a significant proportion of these derogations require national legislators and/or the EU to make use of the leeway offered by the GDPR to implement these in national law. Whether the processing will also be done in a responsible manner depends in particular on the interpretation of the safeguards and measures required under Article 89(1) of the GDPR.

Derogations for research purposes from the information obligation under Article 13 of the GDPR and the right of objection would almost never be allowed if the consultation version of the GDPR Implementation Act were to be adopted in its unaltered form. Such a combination of rights would result in a no-objection system,

applicable to any processing of personal data obtained directly from the data subject. Indeed, the provision of information pursuant to Article 13 of the GDPR requires that data subjects be informed of their rights, including the right of objection, even if their consent does not have to be obtained.

Conclusion

All in all, the GDPR does not seem to obstruct data-intensive health research unnecessarily or disproportionately. Moreover, in a general sense, the GDPR contributes to an overall protection system on the basis of which this research can be carried out in a responsible manner. Nevertheless, the central objective of harmonisation has only been attained to a very limited extent by the EU legislator, as regards the rules specifically governing the processing for research purposes. This lack of harmonisation is reflected in the fact that the GDPR authorises Member States to provide for essential exceptions for scientific research purposes. Such exceptions include the exemption from the consent requirement under Article 9(2)(j) of the GDPR and the derogations from individual rights for research purposes under Article 89(2) of the GDPR. It seems unlikely that such exceptions will be enshrined in EU law in the near future. It is therefore up to the Member States to provide a basis for this in national law, in the absence of which the aforementioned exceptions are not allowed. Furthermore, the conditions and safeguards – to which exceptions for research purposes should be subject – are not sufficiently specified in the GDPR to ensure coherent EU-wide protection. It is therefore highly likely that the negative impact on data-intensive health research due to the lack of legal unity in the EU will persist for the time being.

On the positive side, the GDPR provides sufficient leeway for a data protection regime specifically focused on scientific research. This is achieved by combining exemptions and derogations for research purposes with the requirement to provide appropriate and specific conditions and safeguards. With regard to these safeguards, the GDPR strongly emphasises the importance of institutional governance measures, yet apart from data minimisation, the GDPR contains few concrete indications for the implementation of these measures. This may have adverse effects on the protection of data subjects' rights and interests, as this protection depends to a large extent on the governance measures. In particular, this concerns processing on the basis of broad consent or an exception provision in accordance with Article 9(2)(j) of the GDPR. Further elaboration of these governance measures is therefore a subject that

requires further attention, and should also be taken up by national legislators. This can be done in line with the concrete and specific ethical standards in the recent CIOMS Guidelines or the WMA Declaration of Taipei. It is important, however, that an appropriate choice is made between static legislation and regulations on the one hand and more flexible standards in guidelines on the other. In this respect, the legislator could consider striking a balance by making a code of conduct approved by the Data Protection Authority mandatory for certain processing operations for research purposes. This obligation could be limited to processing for research purposes on the basis of a research exemption from consent in accordance with Article 9(2)(j) of the GDPR. Such an obligation would remove the noncommittal nature of the further elaboration of appropriate and specific measures, all the while maintaining the flexibility of setting standards by means of guidelines arising from the field.

References

1. All of Us Research Program, 2016. Available at https://www.nih.gov/AllofUs-research-program/pmi-cohort-program-announces-new-name-all-us-research-program.
2. BigData@Heart, 2016. Available at https://www.umcutrecht.nl/en/Research/News/UMC-Utrecht-receives-%E2%82%AC20-million-for-big-data-rese
3. Mostert M, Bredenoord AL, Biesaart MCIH, van Delden JJM. Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach. Eur J Hum Genet 2016; 24: 956–960.
4. Consultation version Implementation Act GDPR. Available at https://www.internetconsultatie.nl/uitvoeringswetavg/details.
5. Ploem MC, Essink-Bot ML, Stronks K. Proposed EU data protection regulation is a threat to medical research. BMJ 2013; 346: f3534.
6. Dove ES, Thompson B, Knoppers BM. A step forward for data protection and biomedical research. Lancet (London, England) 2016; 387: 1374–5.
7. The Wellcome Trust. Analysis: Research and the General Data Protection Regulation - 2012/0011(COD), 2016. Available at https://wellcome.ac.uk/sites/default/files/new-data-protection-regulation-key-clauses-wellcome-jul16.pdf.
8. Interdepartementale Commissie Europees Recht. ICER handleiding nationale toepassing EU-Grondrechtenhandvest, 2014. Available at http://www.minbuza.nl/binaries/content/assets/ecer/ecer/import/icer/handleidingen/2014/icer-handleiding-nationale-toetsing-eu-handvest-grondrechten.pdf.
9. ECtHR 16 July 2014, ECLI:CE:ECHR:2014:0716JUD003735909 (Hämäläinen t. Finland).
10. ECtHR, 26 March 1985, App. no. 8978/80 (X en Y v. The Netherlands).
11. E.M.L. Moerel, J.E.J. Prins, Privacy voor de homo digitalis, uit: Preadviezen NJV 2016-1, Wolters Kluwer, 2016, p. 57-58. .
12. The Nuffield Council on Bioethics. The collection, linking and use of data in biomedical research and health care: ethical issues. 2015.
13. Apple's ResearchKit frees medical research, Nature Biotechnology, 2015;33:322. .
14. Article 15d (1) Wet aanvullende bepalingen verwerking persoonsgegevens in de zorg (Wabvpz).
15. M.C. Ploem, Gegeven voor de wetenschap, Regulering van onderzoek met gegevens, lichaamsmateriaal en biobanken, uit: Wetenschappelijk onderzoek in de zorg, Preadvies 2010 Vereniging voor Gezondheidsrecht.
16. Dutch Data Protection Authority, 2016. Available at https://autoriteitpersoonsgegevens.nl/sites/default/files/01_onderzoek_nza-dis.pdf.
17. In contrast to the regulation of medical secrecy in Dutch law, the GDPR only applies to personal data relating to living natural persons. About the gaps in protection of data relating to deceased persons in Dutch law, see: Ploem MC, Dute JCJ. Wetenschappelijk onderzoek na overlijden: goed geregeld? Tijdschrift voor Gezondheidsrecht 2016; 8.

18. Ottes L. Big Data in de zorg | Working Paper | WRR. 2016. Available at https://www.wrr.nl/publicaties/working-papers/2016/04/28/big-data-in-de-zorg.
19. Gymrek M et al, Identifying personal genomes by surname inference, Science 2013; 339: 321-4.
20. Savage N, The Myth of Anonimity, Nature, 2016; 537: 70-2.
21. Rodriguez LL et al. Research ethics. The complexities of genomic identifiability Science 2013; 339: 275-6.
22. Article 29 Working Party. Available at http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.
23. Barocas S, Nissenbaum H. Big data's end run around anonymity and consent, in: Privacy, big data, and the public good: Frameworks for Engagement. Cambridge University Press: Cambridge 2014, pp. 44-75.
24. CJEU 19 October 2016, ECLI:EU:C:2016:779 (Breyer).
25. Quinn P. The Anonymisation of Research Data — A Pyric Victory for Privacy that Should Not Be Pushed Too Hard by the EU Data Protection Framework? European Journal of Health Law, 2016; 24: 1-24.
26. Tene O, Polonetsky J. Privacy in the age of Big Data: a time for big decisions. Stanford Law Rev Online 2012; 64: 63–9.
27. Knoppers BM, Zawati MH, Kirby ES. Sampling populations of humans across the world: ELSI issues, Annu Rev Genomics Hum Genet 2012; 13: 395-413.
28. van Veen EB. Europe and tissue research: a regulatory patchwork, Diagn Histopathol, 2013; 19: 331–6.
29. Lowrance WW. Privacy, Confidentiality, and Health Research, Cambridge University Press: Cambridge, UK, 2012, p. 158.
30. ECtHR 04 December 2008, ECLI:NL:XX:2008:BH1813, § 72.
31. ECtHR 04 December 2008, ECLI:NL:XX:2008:BH1813, § 120.
32. World Medical Association (WMA). Declaration of Taipei on Ethical Considerations Regarding Health Databases and Biobanks. 2016.
33. Council for International Organizations of Medical Sciences (CIOMS). International Ethical Guidelines for Health-related Research Involving Humans. 2016.
34. van Delden JJM, van der Graaf R. Revised CIOMS International Ethical Guidelines for Health-Related Research Involving Humans. JAMA. Published online December 06, 2016. doi:10.1001/jama.2016.18977.
35. European Parliament, Available at http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P7-TA-2014-0212+0+DOC+XML+V0//EN.
36. Council of the EU, Available at http://data.consilium.europa.eu/doc/document/ST-9565-2015-INIT/en/pdf.

37. Explanatory Memorandum, Second House of Representatives, 1997/98, 25 892, nr. 3, p. 126 and 155. Since the government emphasises that the purpose of the GDPR Implementation Act is to approximate existing national law to the greatest possible extent, the Explanatory Memorandum to the former Dutch data protection law is still relevant.
38. Kaye J, Whitley EA, Lund D, Morrison M, Teare H, Melham K. Dynamic consent: a patient interface for twenty-first century research networks. Eur J Hum Genet 2015; 23: 141–6.
39. E.g. Chin JJL, Campbell AV. What – if anything – is special about 'genetic privacy', in: Genetic Privacy: An Evaluation of the Ethical and Legal Landscape, Hackensack, NJ: Imperial College Press, 2013.

Chapter 5

Privacy in Big Data psychiatric and behavioural research: A multiple-case study

Mostert M
Koomen BM
van Delden JJM
Bredenoord AL

International Journal of Law and Psychiatry 2018; 60: 40-4.

Abstract

In Big Data health research, concerns have risen about privacy and data protection. While the ethical and legal discussion about these issues is ongoing, so is research practice. The aim of this qualitative case study is to gain more insight into how these concerns are currently dealt with in practice. For this multiple-case study, the YOUth cohort, a longitudinal cohort focusing on psychosocial development, and Big Data Psychiatry, a pilot study in Big Data analytics on psychiatric health data, were selected. A broad range of relevant documents were collected and semi-structured interviews with stakeholders were conducted. Data were coded, studied and divided into themes during an iterative analytical process. Three themes emerged: abandoning anonymisation, reconfiguring participant control, and the search for guidance and expertise. Overall, the findings show that it takes considerable effort to take privacy and data protection norms into account in a Big Data health research initiative, especially when individual participant level data need to be linked or enriched. By embracing the complexity of the law in an early phase, setbacks could be prevented, the existing flexibility within the law could be utilised, and systems or organisations could be designed and constructed to take relevant rules into account. Our paper illustrates that a close collaboration of experts with different backgrounds within the initiative may be necessary to be able to successfully navigate this process.

Introduction

Big Data is finding its way into health research. Some believe that this will provide unprecedented opportunities for psychiatry.[1] A broad range of issues, however, need to be dealt with. One of the key areas of concern in Big Data health research is related to privacy and data protection,[2] especially when psychiatric or other sensitive health-related data are collected, re-used, linked and analysed.

The rise of such data-intensive health research initiatives has sparked a lively debate about how the use of data should be governed by principles and rules, especially during the adoption of the General Data Protection Regulation (GDPR) in the EU.[3,4,5] Although this debate on normative issues is ongoing, researchers and other stakeholders already need to deal with challenges related to privacy and data protection on a daily basis. They cannot wait until the normative framework is sufficiently crystallized. They are confronted with a level of normative complexity and uncertainty which could have a negative impact, both on achieving scientific goals and on the protection of relevant rights and interests. In the UK, for example, a study has shown that the confusing nature of the regulatory landscape resulted in a culture of caution and (overly) conservative approaches to data sharing.[6]

Against this background, some health research initiatives have attempted to engage with and utilise the potential of Big Data, while at the same time ensuring privacy and data protection. To our knowledge, no qualitative research has been published about how this challenge is dealt with by relevant stakeholders in the specific context of such groundbreaking initiatives. By mapping the relevant challenges faced and solutions sought by those involved in the organisation of such initiatives, valuable lessons can be learned. In this qualitative case study, we analyse two real-world examples of data-intensive psychiatric and/or behavioural research. The study is designed to provide insight into challenges related to privacy and data protection in data-intensive health research, and aims to contribute to a better understanding of how rules and interests can be taken into account in a specific initiative or context.

Methods

A qualitative multiple-case study has been conducted. The case study is a commonly used empirical research methodology, which allows the researcher to investigate a phenomenon in depth and within its real-world context.[7,8] Information was gathered about the Big Data Psychiatry pilot project (hereafter: BDP) and the YOUth cohort

(hereafter: YOUth). This multiple-case study has been evaluated and exempted from further ethical scrutiny by the Research Ethics Committee of the University Medical Center Utrecht. Explicit informed consent has been obtained from all respondents and the management of both initiatives.

Case selection and background

The cases have been selected because of their approaches to different aspects of Big Data research. BDP employs a Big Data approach to its analytical methods, in particular for aiding in hypothesis generation. In YOUth, another aspect of Big Data is reflected in its comprehensive data collection, which is continuously being supplemented and updated. Although no clear and widely accepted definition of Big Data exists, such innovative ways in which data are analysed or captured are considered to be core building blocks of a Big Data approach.[9]

The first case, BDP, aims to explore the potential of Big Data analytics in gaining new insights in the complex psychiatric phenotype. The ultimate goal in BDP is to develop a Big Data analytics instrument that will support health care professionals in their daily practice, for instance by predicting the chance of side effects of medication on the basis of individual patient profiles.[10] A relatively limited set of databases, related to a group of psychiatric patients in Utrecht, was used in the pilot phase of BDP. As a proof of concept, the Cross Industry Standard Process for Interactive Data Mining (CRISP-IDM) was performed on these databases. This resulted in a number of hypotheses and findings, including those related to the themes of aggression during hospitalisation and the effects of medication.[11] Four working groups have been formed in BDP, and one of these working groups is committed to exploring the theme of privacy and confidentiality. This multi-disciplinary working group focuses on how to safeguard the privacy of participants in the pilot phase and the future programme.

The second case, YOUth, is a longitudinal cohort. YOUth aims to explain why some children develop well and others fail to thrive in society by examining how neurocognitive development mediates the influence of biological, child-related and environmental determinants on behavioural development. The cohort study focuses on psychosocial development, ranging from normal development to deviant behaviour and psychiatric disorders. In order to do so, a great variety of health-related data are continuously collected. These data vary from an array of behavioural and cognitive test results to data about environmental, general child and biological factors (including results from EEG and MRI examinations). The YOUth data being

collected will also be linked to other data sources for a broad range of future studies, all in the field of behavioural and psychiatric research.

Data collection

During our data collection phase, both factual information and the views of different stakeholders from the two cases were collected. The factual information includes internal reports of meetings and discussions, research protocols and other documentation, files related to the application for ethical approval, and text on public websites. Our data collection in YOUth took place between February and April 2017, and in BDP between November 2015 and January 2016. The stakeholders were selected on the basis of their variation in backgrounds and involvement in dealing with privacy and data protection related issues related to the cases. Among the stakeholders, the following areas of expertise or backgrounds are represented: management, lead researcher, research staff, privacy and health law, information technology, consultancy, data management, and patient representation. We conducted 14 semi-structured qualitative interviews in total to collect the views of the stakeholders in both cases. The stakeholders were asked questions related to the challenges they experienced regarding privacy and data protection, and how these challenges were dealt with or should be dealt with according to their views.

Data analysis

After collecting data, our research group developed codes and identified themes. The full transcripts and other relevant collected data were coded using NVivo. Mostert and Koomen coded the gathered data. Mostert and/or Bredenoord read the coded data and checked the codes for consistency. During the process of analysis, the codes were adjusted through constant comparison across the transcripts and other relevant data and through discussion within the research group. After reaching consensus on the coding, the themes mentioned below were identified by analysing the data. All interviews were conducted in Dutch and the quotes in the results section have been translated idiomatically. The results were presented to respondents to be checked for accuracy.

Results

During the process of analysis, it became clear that all respondents encountered challenges or issues related to privacy and data protection. After analysis of the

interviews and the other information, three main themes emerged: abandoning anonymisation; reconfiguring participant control, and; the search for guidance and expertise.

Abandoning anonymisation

The first theme concerns the move away from anonymisation as a strategy to prevent the applicability of data protection law. During the first meetings of the working group on privacy and confidentiality in BDP, some of the respondents adjusted their view on what data could be regarded as anonymous. In this phase, the importance of distinguishing between pseudonymous and anonymous data became clear, but the difficulties in making this distinction were also acknowledged:

"(..) the difference between anonymous and pseudonymous data is hard to understand by layman, and it turned out that it is incredibly difficult for jurists to explain what this difference is. Only after this difference has been made clear, you are able to proceed (..)."(R1BDP).

Afterwards, it became clear to all respondents in BDP that irreversible anonymisation according to the standards as set out in the forthcoming GDPR would severely limit the use of data. Another way to proceed had to be found. BDP chose to integrate a Trusted Third Party (TTP) in the data warehouse architecture of BDP. A TTP aims to facilitate the data linkage process on behalf of multiple data holders in a secure way. Only data that are relevant to a certain research question are extracted from local data sources by the TTP. Afterwards, the different personal data sources are linked by the TTP and a unique pseudonym is assigned to the linked data to prevent future data linkage or enrichment on the individual participant level. The TTP was not considered to be a viable solution in YOUth as it would hinder a permanent enrichment of the cohort with external data sources:

"Or a sort of Trusted Third Party, that is always complicated… because than you need to link data for every single research question and that is a barrier to this kind of cohorts. (..) sometimes I just want to enrich my whole dataset (..)." (R11YOUth).

Furthermore, respondents in both cases regarded de-identifying or pseudonymising data as a challenge, especially when it pertained to unstructured or rich data sources, such as open text fields or imaging data. One of the respondents emphasised the difficulties in de-identifying such data as follows:

"Once you start working with big data, (..) you could potentially link data sources to enrich the profile of people in such a way that identification may become very easy.

(..). With a limited number of variables you could already get such unique information that someone could be identified." (R7BDP).

To deal with the above-mentioned challenges, organisational and technical measures were suggested or implemented in both cases. In YOUth, a data access committee was being installed to ensure control over which data would be released, under what conditions and to whom. An important task of this committee would be to determine whether data could be shared without a risk of re-identification. This, however, was considered to be a difficult and time-consuming task. A research data platform was being developed in which this process would be partially automated, while promoting reproducibility and transparency. The value of such a platform was described as follows:

"At the moment, we are building a research data platform (..) in which we will combine our expertise to store the data prepared for release and which allows the data manager to easily assess whether a combination of data and variables may lead to re-identification. And the system will (..) prevent that this combination of data will be released, (..) And the second advantage is that you will register everything that is done with the data." (R12YOUth).

An alternative to sharing the data itself, mentioned by respondents from both cases, is to only release data analyses. In this way, the data would remain local and only analyses, which would bear no risk of re-identification, would be shared with third parties. Respondents noted that such a system solves many privacy-related problems, but also limits potential data use. One respondent in YOUth, for instance, suggested that this approach would exclude the permanent enrichment of one cohort with the data from another cohort, because this would require access to the raw data.

Reconfiguring participant control

The second and most discussed theme concerns the challenge of allowing participants to control the use of their data. In BDP, most respondents agreed that it is important to obtain informed consent from participants for the re-use and linkage of their personal data. At the same time, some of these respondents recognised the disadvantages of this approach, such as a possible lack of inclusion of data from patients who are already underrepresented in health research and the risk of consent bias. In meetings of the working group on privacy and confidentiality in BDP, participants discussed the fact that it is not always required to obtain informed consent for research on personal data, according to Dutch law, and that there could be other ways to allow participants to exercise control. Furthermore, the importance of making participants

aware of the use of their psychiatric data was stressed by some of the respondents in BDP, as the following quote illustrates:
"There is something special about psychiatric data. (..) some people may not agree, but a diagnosis is not always in your best interest. (..) it is in particular sensitive data because it is about your mental well-being, your mental state, with all kinds of possibilities and impossibilities in the work sphere. It therefore is, in short, very private and sensitive information. When you will link such data on the individual level, it becomes relevant whether the participant is aware of this." (R5BDP).

A question raised in both cases was how to obtain informed consent from participants when their personal data would be used for a broad range of purposes and would be linked to other data sources. In YOUth, multiple respondents described that concerns among the participants arose because they were explicitly asked for consent to request their data from several other databases. These respondents found it hard to eliminate these concerns among their participants related to future data linkages with external data sources.

Another topic of discussion was what should be done to allow participants to control the use of their data in the long-term. In YOUth, researchers especially struggled with the question whether re-consent should be obtained from children once they reached adulthood. The Research Ethics Committee (REC) recommended actively offering children the possibility of opting-out as soon as they would become legally competent. A suggestion of multiple respondents in BDP was that, in the future, all patients should be enabled to be informed and adjust their preferences in an online environment, such as a patient portal. One of the respondents underlined the importance of utilising digital tools as follows:
"I think that such a tool is a beautiful way to not take away the control from participants. That you will not ask for consent once and open the floodgates. And I think that it is a good way to ensure the trust of people in such a whole project, because that is what it is all about." (R5BDP).

Other respondents in BPD, however, pointed out that while developments like dynamic informed consent procedures and patient empowerment are laudable, they could also threaten research, mainly because of the risk of (consent) bias. Moreover, some respondents stressed that the availability or accessibility of appropriate tools to implement dynamic consent procedures was lacking.

Interesting in this context is the fact that multiple respondents emphasised that safeguards and measures other than informed consent are just as important when it comes to ensuring the trust of participants and/or the public. The active

participation of participants was regarded as an important component of these measures. In both cases, they were still in the phase of exploring ways to implement patient participation in their project. In YOUth, a parent representative panel was already established and embedded in the organisation. One of the respondents emphasised the importance of such measures as follows:
"Constantly involving them [participants] in what we are doing. Is this possible when you view it from a patient perspective? Is it right? Is there support for doing it this way?" (R6BDP).

Other suggested measures include the clear designation of responsibilities and tasks, implementing accountability and oversight mechanisms, and certification of the security measures taken. A measure that received particular attention in YOUth was the implementation of policy on the return of clinically relevant incidental findings, after the REC urged YOUth to elaborate on what to do with this kind of findings.

The search for guidance and expertise

The third theme concerns the search for guidance and expertise, which was regarded as an issue by the majority of respondents in both cases. Respondents described that they often struggled with uncertainties, mainly about legal norms, when setting up the project or cohort. To deal with uncertainties in the field of privacy and confidentiality, a working group was established in BDP in an early phase. Related to the feedback and advice of this working group, a respondent noted:
"It still really is a matter of pioneering and finding out... It really surprises me that all those members of the privacy group say: you are really front-runners and it is very praiseworthy that you are carefully addressing this. And at the same time, it is by far not good enough. Then I think, wow, if this is the case, than it tells you something about the state of affairs, also in other areas. (...).
And we try to be the most virtuous of them all (..). As a result, we literally have been delayed several times." (R1BDP).

To avoid such delays or other setbacks, some respondents underlined that it is essential to take into account the legal and ethical aspects at the very beginning of the project. The need for collaboration between different areas of expertise, such as legal, security and ICT was also emphasised. Furthermore, many respondents expressed the need for more specific, uniform and up to date guidelines and best practices for data linkage, on the national and/or international level. Some respondents suggested that a national advisory body, where all the needed expertise and guidance is concentrated, would be of great value. In YOUth, this need for more specific and

uniform guidance was especially focussed on how to obtain consent for future data linkage, as this quote illustrates:
"Well, a dilemma I encounter is that there are or were few good examples, and I still wonder if we are doing it right, how to obtain consent for that linkage. (..). Also because I think that every institution has different requirements and that is what makes it so complicated, because you really want one golden standard for how it should be asked." (R11YOUth).

Some respondents explained why it was so hard to provide clear and uniform guidance on privacy aspects. A first factor is that care and research are closely related, and that they are getting more and more intertwined. Secondly, respondents pointed out that there are many (upcoming) changes in laws and regulations. In addition, the existing national code of conduct related to the use of personal data in health research was not updated for years. The consequences of these developments, and the intention of the legislative bodies, were understood as follows by one of the respondents:
"It meant that everybody had to guess what the best approach would be. And what we saw is that the legislator perhaps consciously created a grey area to be sure that people would keep thinking and to prevent fixed rules. This means it is not that strange that if you talk to people, who do research or are busy with patient care, that they do not really understand." (R7BDP).

Discussion

By studying the two cases, we aim to provide insight in how challenges related to privacy and data protection are dealt with in real world examples of data-intensive psychiatric and behavioural health research. A first insight is that anonymisation was regarded largely impracticable in both cases, especially when data sources needed to be linked or enriched. Organisational and technical measures have been implemented with the aim of complying with the law and mitigating risks, which most notably resulted in the use of a TTP and the development of a research data platform to access and link data. Secondly, it becomes clear that the search for meaningful and proportionate ways to allow individuals to control and be aware of the use of their data is ongoing. This aspect is considered of great importance by some, especially when it comes to linkage of personal data related to psychiatric disorders. Improvements of the (one-off) broad informed consent procedure are being considered, in particular by means of utilising digital tools to ensure a more

ongoing engagement with participants. Thirdly, uncertainty about how to comply with the law is perceived as having negative impact on the initiatives. This uncertainty is mainly attributed by respondents to a lack of easily accessible expertise and guidance, but also to the recent changes in data protection law.

An issue that connects all the described themes and many of the findings is the struggle with legal complexity. Respondents reported a broad range of negative consequences related to legal complexity, such as uncertainty, delays and other setbacks. This critique of the law does not seem to be unique to these cases. Laws that govern the use of data for research, and in particular data protection laws, are often reported to be confusing, open to varying interpretation or burdensome.[6,12] However, what needs to be taken into account is that broad or open norms also have advantages. Without open norms, the law would be static and inflexible. This would result in major problems, since the multifaceted and continuously evolving data-intensive health research landscape requires a considerable degree of flexibility. Paradoxically, it has been pointed out that because of the complex regulatory landscape, the existing flexibility within the legal framework to address some of the regulatory hurdles is often overlooked in practice.[6]

In the cases, multiple ways forward were suggested that could help mitigating legal complexity. One of these suggestions is the drawing up of context-specific guidelines for data linkage. Currently, some initiatives like BBMRI-ERIC are already working towards official approval under the GDPR of an international code of conduct for personal data processing in health research.[13] Although a context-specific code of conduct could indeed help reduce legal uncertainty, it is unrealistic to expect that this will mitigate the need to deal with legal complexity on the level of the initiative. Open and often fluid norms will remain, and these norms will need to be interpreted and translated into concrete, effective and proportionate rules and measures on the level of each single data initiative.

From the cases, we can learn how this challenge of taking the complex interplay of norms and practical requirements into account could be approached. A first key element of the approach is to identify and address privacy-related issues in an early phase. In BDP, the identification of privacy-related issues has in particular been done by establishing a multi-disciplinary working group on privacy and confidentiality during the beginning of the pilot phase. The meetings of this working group inspired multiple important decisions made and measures taken in BDP, including the move away from anonymisation as the main compliance strategy. Afterwards, discussions within BDP focussed on finding another way to link multiple personal data sources

while respecting relevant rules on privacy and data protection. In YOUth, the search for a way to link personal data sources to enrich the cohort in a responsible and effective way also emerged during the start-up phase. This brings us to a second key element of the approach in both cases, which can be labelled as privacy by design. The privacy by design approach is a form of value-sensitive design, which is characterised by the embedding of privacy-enhancing and preserving measures directly into the design and operation of systems, processes and organisations.[14] The use of a TTP in BDP and the development of a research data platform in both cases are measures that resemble this approach. Other examples of research data platforms have been described in the literature, with the initiative called DataSHIELD serving as a prime example. The DataSHIELD platform demonstrates how a value-sensitive design could contribute to linking and utilising multiple data sources for health research purposes, while respecting relevant rights and values.[15,16,17] Another incentive to embrace a value-sensitive design approach is provided by the GDPR. In the GDPR, a novel legal obligation emerged under the title 'Data protection by design and by default'. This obligation is subject to high fines and considered to be among the most innovative and ambitious norms of the GDPR.[18] Nevertheless, it has proven very difficult to encode data protection principles and rules in systems for data processing in health research. The idea of integrating legal norms in information processing systems during the design phase is by some regarded to be "at odds with the dynamic and fluid nature of many legal norms".[19] It therefore seems that both the potential and the limits of such value-sensitive design approaches need to be recognised in practice and deserve further study.

Finally, some of the measures taken in the cases illustrate that compliance with the law was not regarded sufficient to guarantee that all relevant rights, interests and values would be taken into account. In YOUth, for instance, the presence of a policy on the return of clinically relevant unsolicited findings was considered to be an essential measure to respect participant's interests. Another example was the ambition in both cases to involve and engage with participants in a meaningful way, and on an ongoing basis. These (and other) measures go beyond what is required by the law, but may be necessary to meet the expectations of participants and society regarding the conduct and activities in data-intensive health research. What the consequences of a failure to meet these expectations might be is illustrated by the disturbing developments in the care.data program.[20] It should therefore be prevented that the focus on compliance with the law completely overshadows the protection

of other relevant rights and interests and values, just because they are not that well operationalised in the legal system.

This study has some limitations. First of all, only two cases were analysed and a limited number of interviews took place, so that the findings do not necessarily represent the most common or relevant privacy and data protection issues. However, we do feel that within these two cases a sufficiently complete image was extracted due to the fact that most of the individuals who actually dealt with relevant issues were interviewed and a broad range of available documents were studied.

Concluding remarks

Overall, the findings show that it takes considerable effort to take privacy and data protection norms into account in a data-intensive health research initiative, especially when individual level data need to be linked or enriched. By embracing the complexity of the law in the initiative in an early phase, setbacks can be prevented, the existing flexibility within the law can be utilised where appropriate, and systems or organisations can be designed and developed so that they take relevant rules into account. The cases and discussion illustrate that a close collaboration of experts with different backgrounds within the initiative may be necessary to be able to successfully navigate this process.

References

1. Monteith S, Glenn T, Geddes J, Bauer M. Big data are coming to psychiatry: A general introduction. Int J Bipolar Disord 2015; 3: 21.
2. Mittelstadt, BD, Floridi L. The ethics of big data: Current and foreseeable issues in biomedical contexts. Science and Engineering Ethics 2016; 22: 303–41.
3. Mostert M, Bredenoord AL, Biesaart MCIH, Van Delden JJM. Big data in medical research and EU data protection law: Challenges to the consent or anonymise approach. European Journal of Human Genetics 2016; 24: 956–60.
4. Ploem MC, Essink-Bot ML, Stronks K. Proposed EU data protection regulation is a threat to medical research. BMJ 2013; 346: f3534.
5. Sethi N. Reimagining regulatory approaches: On the essential role of principles in health research. SCRIPTed 2015; 12: 91–116.
6. Sethi N, Laurie GT. Delivering proportionate governance in the era of eHealth. Med Law Int 2013; 13: 168–204.
7. Baxter P, Jack S. Qualitative case study methodology: Study design and implementation for novice researchers. The Qualitative Report 2008; 13: 544–59.
8. Yin RK. Getting started: How to know whether and when to use the case study as a research method. In: Yin RK (ed.), Case study research: Design and methods (pp. 3–25). CA, USA: Thousand Oaks, 2014.
9. Mayer-Schönberger V. Ingelsson E. Big data and medicine: A big deal? Journal of Internal Medicine 2018; 283: 418–29.
10. Scheepers FE, Menger V, Hagoort K. Data science in psychiatry. Tijdschrift voor Psychiatrie 2018; 60: 205–9.
11. Menger V, Spruit M, Hagoort K, Scheepers FE. Transitioning to a data driven mental health practice: Collaborative expert sessions for knowledge and hypothesis finding. Computational and Mathematical Methods in Medicine 2016; 9089321.
12. Koops BJ. The evolution of privacy law and policy in the Netherlands. Journal of Comparative Policy Analysis 2011; 13: 165–79.
13. BBMRI-ERIC. A Code of Conduct for Health Research, 2017. Available at http://code-of-conduct-forhealth-research.eu.
14. Cavoukian A. Privacy by design: The definitive workshop. A foreword by Ann Cavoukian, Ph.D. Identity in the Information Society 2010; 3: 247–51.
15. Budin-Ljøsne I, Burton P, Isaeva J et al. DataSHIELD: An ethically robust solution to multiple-site individual-level data analysis. Public Health Genomics 2015; 18: 87–96.
16. Wallace SE, Gaye A, Shoush O, Burton PR. Protecting personal data in epidemiological research: DataSHIELD and UK law. Public Health Genomics 2014; 17: 149–57.
17. Wolfson M, Wallace SE, Masca N et al. DataSHIELD: Resolving a conflict in contemporary bioscience—Performing a pooled analysis of individual-level data without sharing the data. International Journal of Epidemiology 2010; 39: 1372–82.

18. Bygrave LA. Data protection by design and by default: Deciphering the EU's legislative requirements. Oslo Law Review 2017; 4: 105–20.
19. Koops BJ, Leenes RE. Privacy regulation cannot be hardcoded. A critical comment on the 'privacy by design' provision in data-protection law. International Review of Law, Computers & Technology 2014; 28: 159–71.
20. Carter P. Laurie GT, Dixon-Woods M. The social licence for research: Why care.data ran into trouble. Journal of Medical Ethics 2015; 41: 404–9.

Chapter 6

Responsible data sharing in international health research: A review of principles and norms

Kalkman S
Mostert M
van Thiel GJMW
van Delden JJM

Submitted.

Abstract

Aim: To identify a coherent set of ethical principles and norms to govern responsible data sharing for international health research.
Methods and results: We performed a review of ethical guidelines, policy documents and literature sources for ethical principles and norms pertaining to data sharing for international health research. We observed an abundance of principles and norms with considerable convergence at the aggregate level of four overarching themes: societal benefits and value; distribution of risks, benefits and burdens; respect for individuals and groups; and public trust and engagement. However, at the level of principles and norms we identified substantial variation in the phrasing and level of detail, the number and content of norms considered necessary to protect a principle, and sometimes even contradiction between norms.
Conclusion: Though providing some helpful leads for further work on a coherent governance framework for data sharing, the current collection of norms and principles is still too haphazard, non-uniform and sometimes even contradictory to serve as sufficient guidance in itself. Our work highlights the need for considerable investments and expertise to further develop and implement a governance framework for international data sharing projects.

Introduction

Recently, a number of multi-stakeholder initiatives have been funded to develop data-driven translational research platforms to improve patient outcomes and reduce the societal burden of specific disease areas in the European Union (EU).[1,2] The Innovative Medicines Initiative's (IMI) BigData@Heart is an example of a consortium that is currently designing an international data sharing platform to stimulate drug development and personalised medicine for cardiovascular disease. To ensure responsible use of data in BigData@Heart as well as similar research projects, good governance of data sharing and data access is critical.[1]

So far, no blueprint of a broadly accepted governance framework exists. The recently adopted General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679) will not be able to provide for the necessary guidance in full, since specific provisions for scientific research may still be formulated at the level of national jurisdictions within the EU.[3] Moreover, compliance with the law does not always guarantee that data is used in morally acceptable ways, or that public trust is secured.[4] The evolving landscape of big health data raises new questions about both familiar ethical concepts (such as privacy, confidentiality and informed consent), as well as novel ones. These developments indicate that innovative and adaptable governance models are highly needed to establish a practice of truly responsible data sharing.

To identify what elements are considered inherent to a governance structure for responsible data sharing within (consortium-wide) platforms for international health research, we reviewed frameworks for data sharing as described in ethical guidelines and the academic literature. This study was driven by the question: What are the ethically relevant principles and norms so far developed by (international) working groups or professional organisations with respect to international data sharing in health research?

Methods

Search and selection

We performed a review of principles and norms for responsible health data sharing, as developed in guidelines, policy documents and the academic literature. National and EU laws and regulations were excluded from this study because we were primarily interested in elements of a governance framework that provides comprehensive moral guidance, not only enforces legal compliance. Even though the law does require

the implementation of a number of organisational and technical measures, what a governance framework exactly looks like is ultimately to be developed in practice.[5]

Relevant guidelines and policy documents on data sharing in international health research were identified with help from academic and industry consortium partners with expertise in health law, regulatory science and research ethics. Relevant literature was identified through a systematic search in four academic databases (See Appendix I for a breakdown of search terms). Search strings were adjusted to the type of database to restrict superfluous results to a minimum (See Appendix II). For inclusion, publications were required to present a coherent set of principles and/or norms that could potentially function as or at least be construed as part of a model or framework for responsible data sharing. Documents were included if the content was developed with the purpose to inform policy decision-making and preferably by or in collaboration with (international) working groups or professional organisations active in the field of health data sharing.

Publications that were limited to a discussion of benefits, imperatives or challenges for health data sharing or IT infrastructures for Big Data research were deemed not relevant to the purpose of this review. All sources that were not of relevance to the European context were also excluded (e.g., practice guidelines for low and middle income countries). Since we were also specifically interested in developments over recent years, we limited our search to sources published between 2006 and 2017. Only sources published in English were eligible.

Data extraction and analysis

From all included guidelines and references we extracted the following data: author names, year of publication, organisation or working group, countries the recommendations apply to (EU/US/international), and the status of the recommendation. By 'status' we mean whether the recommendation, for example, has a legal basis, is an ethical guideline, comprises lessons learned, or is an academic proposal. Textual analysis of sources for principles and norms was performed by two independent assessors using the Covidence online support tool for systematic reviews and NVivo qualitative data analysis software (QSR International, Version 11).

Results

Selection and data extraction

The expert consultation resulted in the inclusion of 10 ethical guidelines by 7 different organisations or working groups (Table 1). The selected guidelines were published between 2007 and 2017. The literature database searches resulted in a total of 892 unique records. Ultimately, we included 27 articles for final review (Figure 1). Identified principles and norms were grouped in themes as a means to structure the research findings. Descriptive themes were established through an iterative method and with consensus of all study authors.

Themes, principles and norms

Following data extraction from all sources, the identified principles (and the respective norms promoting those principles) could be grouped among four overarching themes: (1) *Societal benefits and value*; (2) *distribution of risks, benefits and burdens*; (3) *respect for individuals and groups*; and (4) *public trust and engagement* (Table 2).

Societal benefits and value

In most sources, data sharing activities were required to be governed by principles that overall maximise health benefits or wellbeing (both public and individual) and that serve ends of social value. To realise the potential benefits, sources underpin the importance of the *quality of the data* to be shared, and the *scientific validity* and *social value* of the study protocols submitted by researchers in order to use the data. Once quality and validity have been established, many sources demand a data sharing infrastructure that is *accessible*, enables *efficient* use, is highly *interoperable* and *sustainable* for the future (See Table 2).

In terms of how to bring the principles into practice, sources rely on a wide range of norms, rules and recommendations. First, sources deduce from the potential benefits that there in fact exists a *duty to share* data for scientific research.[6] To effectuate *the duty to share*, sources state that awareness about the benefits of data sharing should be raised among stakeholders, and that collaborative partnerships and data sharing practices should be promoted.[7] Medical journal editors and industry associations have come forward with statements about researchers' and companies' duty to share their clinical trial data.[6] Other recommendations include devoting efforts and resources to alleviate disincentives for data sharing, such as publication moratoria.[8] The sharing of well-managed datasets and commitments to *disseminate*

the results generated from the data (mostly through reports and supporting scientific publications) are considered an equally important element of maximising benefit of data sharing.[9–11]

Continuous efforts should be undertaken to improve and maintain *data quality and reproducibility.*[12,13] Demands with respect to data management and curation include cooperatively developing and implementing quality standards or quality threshold metrics that are submitted to continuous renewal and improvement.[14–17] Sources emphasise the need for data control, compliance with quality standards and feedback mechanisms[7,15] at every stage of data processing.[16] The use of central repositories is recommended for deposition of data.[17] To maximise *scientific and social value*, data access requests will need to be submitted by qualified researchers who are able to justify the research purposes,[18–21] and attest to the use of rigorous scientific methods.[22,23] Those providing access for secondary use should in turn secure comprehensiveness of the data.[17]

Accessibility of the data is considered a shared responsibility of researchers, sponsors and research ethics committees. These actors should (deliver reasonable efforts to) maximise accessibility, and encourage each other to do so too.[7,9,13,14,16,19] Accessibility is further enhanced through harmonisation of data access conditions and procedures,[10] and by communicating these to stakeholders.[7,15,24] One source states that access to data should be granted at the lowest possible cost to the international research community.[9] Many sources consider the development of strategies, processes and/or systems that help secure long-term accessibility (e.g., through funding) and *sustainability* of the organisation of great importance.[7,15,16,20–22,25,26] It should be made clear how the data will be dealt with in the event of discontinuation of the data holder,[14,16] or a change of ownership.[27] Uniform policy is required with respect to the duration of storage,[16] and the disposal and destruction of data.[27]

Interoperability is enhanced by cataloguing data in a consistent manner,[11] according to internationally accepted standards and norms,[7,15,16] by incorporating standardised design elements that provide for compatibility,[16] and through harmonisation of regulatory frameworks for data sharing in Europe.[10] Documentation of data quality and origin should be readily available, verifiable,[16] accurate, unbiased and proportionate.[7] For those who have been granted access to data, validation exercises should be allowed whenever possible.[15]

Distribution of risks, benefits and burdens

Many sources require that the burdens and benefits of data sharing are fairly allocated. In other words, data sharing efforts should adhere to principles of distributive justice (See Table 2). *Benefit sharing* and *reciprocity* is distinguished between participants and researchers, as well as between researchers, secondary users, communities and funders.[9,19,28] One source states that it should be assured that benefits are shared "as broadly as possible"[16], especially when data is collected from vulnerable communities.[27] *Equitable access* is ensured by fair access fees and transparency rules.[16,19] Commercial interest is generally not considered a reason to restrict access to data. However, access should be based on balanced arrangements between public and private parties.[15]

Sources also emphasise the need for establishing adequate systems for *recognition* and *attribution*, that are designed in such a way that due credit and acknowledgment is given to all who contributed to the results.[7,29] These principles are promoted through the application of intellectual property (IP) laws to data access arrangements.[15,27,30] In general, policy should make sure to cover benefit sharing and IP issues as transparently as possible, and for it to be communicated appropriately.[16,27] Researchers are required to report back to the relevant data holders a list of publications and patent issues arising from the database's resources.[16]

Respect for individuals and groups

Respect for persons and the duty to minimise risks in data sharing efforts is represented by an abundance of identified principles, norms, rules and recommendations (See Table 2). From the principle of *respect for autonomy* it follows that the purposes to which data is shared should be consistent with the (scope of the original) informed consent.[14,16,22,23,29] Some sources differentiate between 'specific informed consent' and 'broad informed consent' for a range of future data uses.[13,14,16,27] When future use is specified at the time of data collection, or the data are collected for a given research project, specific informed consent from individuals is required.[14,27] When this is not the case, some sources permit the conditional use of broad consent models.[13,14,16,19,27] Valid broad informed consent relies on certain (additional) safeguards, such as a proper governance framework and the provision of sufficient information to participants.[13,14,16,27] During and/or after the informed consent process, sources state that participants should be informed about the topics as listed in Table 3. Clear and easy-to-use processes should remove barriers for participants to withdraw their consent for the use of their data at any time.[14,16,19,27] *Rights* that are considered

relevant for participants are listed in Table 3. Furthermore, policies and procedures are recommended for when and how to re-contact participants,[13,14,16,27,31] in particular with respect to the return of unsolicited findings, and how participants can request access to their data.[12,16]

If informed consent for data access cannot reasonably be obtained ("impossible" or "impracticable"), waivers of informed consent may potentially be issued.[14,16,27,32,33] Some of the sources state that waivers of informed consent for data (re-)use should be issued after approval of a research ethics committee (REC) only, and "in accordance with applicable law" and "ethical principles".[16,34] The Declaration of Taipei restricts waivers to the event of a "clearly identified, serious and immediate threat (..) to protect the health of the population",[27] while the Council for International Organizations of Medical Sciences (CIOMS) guidelines demand that the study has important social value and poses "no more than minimal risks".[14] An alternative is to have RECs allow the conditional use of an 'informed opt-out' procedure.[14] Even in cases where no express consent has been given, however, individuals should be able to express preferences regarding the use of their data—at least to the extent practicable.[19]

Norms that help protect *privacy* and *confidentiality* include the establishment and periodical updating of security measures, protocols and other protective safeguards,[12,13,15,16,18,19,27,35] which are proportionate to the use and nature of the data.[7,32] Substantial support was observed among sources for the requirement to only store and share data that is de-identified (anonymised or coded).[16,18,36,37] At the same time, the limits of anonymity and confidentiality are acknowledged and should be anticipated.[14] [16] One source states that use of anonymised data should generally be avoided because it makes it impossible to add patient-level data and/or to re-contact participants.[29] In all cases, researchers are said to have the obligation to inform individuals that complete confidentiality can never be guaranteed.[29] There is agreement among sources on the rule that the sharing of identifiable data or permission for re-identification should only be allowed for research purposes (unless ordered by law) and after approval "conform applicable procedures".[16,19] Terms include access limitation to those with a need-to-know,[18] and restrictions on who may have (third party) access to (potentially) identifiable data.[12,14,16]

Data security is further enhanced if technical alternatives for physical transfer of data are explored, such as the use of secure data access centres and remote data access facilities. [19,29] To prevent unauthorised access or any other misuse, robust infrastructures will need to arrange for identity verification and authentication before access is granted.[16,18,19] Infrastructures should also monitor and document any access

to identifiable data,[16] and implement feedback mechanisms for data security.[7] Policy should include statements about how confidentiality is practically maintained,[14] and that users must refrain from any attempt to (re-)identify participants.[7,13] Essential to secured sharing is education and training of researchers on issues such as data security and privacy compliance.[11,38]

Public trust and engagement

Many sources report on principles and norms that relate to maintaining *public trust* and engaging in *public and patient involvement* and/or *participation.* Public trust and engagement constitute a theme that has instrumental value to maximise benefits, promote respect for persons, minimise harms *and* to protects principles of social justice. Nevertheless, we treat public trust and engagement as a separate moral category to illustrate the emphasis that it has been given in the reviewed sources.[11,38] Key principles reported by the sources that foster public trust and engagement are shown in Table 2.

Overall, sources emphasise the need to develop formats and mechanisms that enable effective deliberation with relevant stakeholders—including participants, the public, funders and the research community—about important issues of data sharing.[7,10,13,14,16,22,24] More specifically, participation should be increased in the design, governance and review of data initiatives—of which the results should eventually translate into policy. Preferably, a regular process of reviewing and modifying data access policies, protocols and procedures should be in place,[15,16] which pays heed to relevant issues that may change over time (e.g., IT, legal and/or cultural issues).[15] Other opportunities for patient and public involvement include events and workshops to disseminate research findings, as well as organising lay presentations on panels, steering committees and working groups.[11,18]

The principle of *transparency* can be brought into practice through different mechanisms. First and foremost, transparency needs to exist in all workflow of data sharing activities and transactions (including documentation).[12,20,25,26,39] Especially transparency in data sharing transactions is flagged as an essential component of responsible data sharing.[40] The principle is also effectuated through the dissemination of public information about ongoing data sharing activities.[38] Items that are proposed to be included in such public information are listed in Table 3. At the same time, researchers and institutions will need to raise awareness and increase understanding among the public towards the need for data sharing to democratise health research.[18,25,38]

Special consideration was given to the importance of effective governance systems as a means to promote *integrity, solidarity* and *accountability* in data sharing activities.[12–14,19,38,41] Each international collaborative data research initiative is expected to operate "within an explicit public ethics and governance framework".[13] The governance structure should clearly outline the *responsibilities* of designated individuals or entities,[18] establish measures for accountability (e.g., whether secondary use has met the intended purposes and sanctions for breaches),[18] and install mechanisms for monitoring, audits and general oversight (e.g., good stewardship of stored data).[13,14,16,18,19] A more specific recommendation is to establish a governance committee to oversee policy developments.[8] Compliance with existing legal requirements, ethical principles and collaborative agreements is considered paramount.[16,18,21,22,34] Particularly, investments need to be made in fostering *professionalism*—which involves education and training of professionals and other staff—and communication with participants and the public.[11,16,38] *Social accountability* arises from engagement of individuals in society, supported by organisations that communicate to individuals and society about the expectations and failures of data governance.[13]

In most sources, review and approval procedures by an independent REC (or comparable review body) play an important part in discussions about responsible data sharing for health research.[11,34,42] Some sources state that an REC (or comparable body) must review and approve every study using collected data.[14,16,27,34] Some aspects of REC review have already been discussed in the context of respect for individuals and groups. The full list of items or situations that are considered subject to ethics review and approval can be found in Table 4. Data access should be based on the legitimacy of the research purpose,[23] objective and clearly articulated criteria (as recorded in policy documents),[21] and restricted to researchers who have received adequate data security training,[11] and who are subject to institutional oversight and effective sanctioning.[13,16,18] When access to data is granted, agreements should specify the terms of access.[16,19,27] Transactions can be responsibly facilitated through the use of binding data access agreements (DAAs), such as data transfer agreements (DTAs).[6,25,28,31,37] Ideally, these DAAs follow a standardised format to regulate access uniformly and consistently. DAAs should include arrangements to promote good practices to enable quality control,[15] arrangements for a secure transfer,[19] and appropriate and effective means to sanction non-compliance.[19]

Discussion

This systematic review of the academic literature and research guidelines provides a unique overview of ethical principles and norms that are considered inherent to a governance framework for responsible data sharing. Fourteen guidelines and 27 international academic publications were qualitatively analysed. We observed an abundance of principles and norms with considerable convergence at the aggregate level of four overarching themes: societal benefits and value; distribution of risks, benefits and burdens; respect for individuals and groups; and public trust and engagement.

In terms of societal benefits and value, it is considered necessary by some to raise awareness about the duty to share health data, and to secure that only high-quality data is shared for scientifically valid proposals. Systems for data sharing should allow for efficient use, and be highly interoperable and accessible, as well as sustainable for the future. To ensure fair distribution of risks, benefits and burdens, effective mechanisms for benefit sharing will need to be in place. Collective evidence generation requires governance that has systems for recognitions, attribution and ownership built in. Respect for individuals and groups covered a range of identified norms and principles, among which the principles to respect privacy and confidentiality were by far the most prominent. There is a growing consensus that absolute anonymity or confidentiality cannot be guaranteed, despite the common requirement to de-identify data to maintain confidentiality/privacy. Moreover, because of the nature of data sharing activities, it is acknowledged that alternatives will need to be devised for traditional, specific informed consent. What is more, it is recommended in most of the sources that an ethics committee (or a comparable body) reviews and approves data access requests. Lastly, public trust is crucial to responsible data sharing. In this relation, accountability, transparency, integrity and professionalism are key principles. Continued stakeholder engagement, from study design to the dissemination of research findings, can and should be facilitated using different methods. The themes we have identified share considerable similarities with the moral considerations of a framework for public health ethics.[43] This suggest that the ethics of international data sharing is probably best captured by moral duties that arise from the interactions and relationships between health care professionals, various public and private actors and the public. We hasten to mention that our thematic categorization is not intended as a new framework in itself. Rather, our

thematisation helps to identify common grounds and to structure various norms and principles.

At the level of principles and norms we observed substantial variation in: (1) the phrasing and level of detail of norms and principles, (2) the number and content of norms considered necessary to protect a principle, and sometimes even (3) outright contradiction between norms. An example of (1) is that some sources reported only in very general terms on relevant principles (e.g., "data sharing should be transparent" or "access should be ensured"), while others provided more detailed descriptions (e.g., "the public should be continuously updated about ongoing data sharing activities" or "ensure low data access fees"). Point (2) is exemplified by the diversity of norms related to informed consent and exemptions from (specific) consent requirements. Only some of the sources explicitly allow the conditional use of broad informed consent models or opt-out procedures. With respect to point (3), while one source would discourage the use of anonymised data other sources would actually demand complete de-identification. Though the identified principles and norms provide some helpful guidance on an impressive range of items, these three points indicate that the identified collection of (proposed) principles and norms, in its current state, is too varied and non-uniform to be used as international guidance for data sharing activities. Add to this the already extensive variation in national laws and regulations, and the need for a harmonised and comprehensive governance framework becomes evident.

Principles and norms relevant to responsible data access and sharing will thus need to be embedded in an international governance framework that is adaptable to local or specific issues. This review has done some of the work by identifying those principles and norms. The next step is to feed the principles and norms into coherent and practical guidance for stakeholders. Although different collaborative partnerships have already taken the initiative hereto,[44,45] we stress the need for considerable investments and expertise to actually further develop and implement a governance framework for international data sharing projects. Substantial efforts of an interdisciplinary team are essential to take legal, ethical and practical requirements into account.[5]

A particular issue of importance we wish to address here is the sparse guidance on how to effectively deal with the limitations to preserve anonymity and confidentiality of shared data. One fairly undisputed recommendation is to inform participants about the limits of anonymity and confidentiality. However, the extent to which the other recommendations apply to the use of anonymous data largely remains unclear.

For example, will access to anonymous data always need to be subjected to review and approval by an ethics committee? None of the sources explicitly states that the access to and use of *anonymous* data should be subject to ethical review or other accountability measures. This is in spite of the fact that de-identification is, on its own, not regarded as a safe strategy for ensuring that the rights and interests of participants are protected.[13] We therefore believe that—in order to truly safeguard the rights and interests of participants—future work should concentrate on the development of measures to establish public trust in data sharing activities, at all levels of (de-)identification. This review has already identified a number of principles and norms to help establish public trust (such as transparency and accountability). We recommend that a governance framework thus goes beyond 1) simply acknowledging the limits of anonymity and/or 2) requiring de-identification at all costs (at the expense of data quality), and that the key to resolving limitations in anonymity lies in the explicit connection with public trust.

The themes we have identified share considerable similarities with the moral considerations of a framework for public health ethics.[46] This suggest that the ethics of international data sharing is probably best captured by moral duties that arise from the interactions and relationships between health care professionals, various public and private actors and the public. We hasten to mention that our thematic categorization is not intended as a new governance framework in itself. Rather, our thematisation helps to identify common grounds and to structure various norms and principles in such a way that the basic structure of a governance framework becomes visible. We acknowledge that certain principles could be categorized as belonging to more than one than theme, and norms and recommendations as serving more than one principle. This review was also limited to expert-selected guidelines and a selection of peer-reviewed literature on the topic of data sharing for health research. We are aware that our findings, particularly the body of sources identified by experts, cannot make any claims to comprehensiveness. A plethora of policy statements on data access and data sharing exists at the level of governmental bodies, regulatory agencies, and public and private institutions. Specifically, pharmaceutical companies have been increasingly active when it comes to the development of policy on data transparency.[47] Also the European Medicines Agency's policy on publication of clinical data deserves special mentioning in this context.[48]

For this study we aimed to capture what norms and principles have been expressed by (international) collaborative working groups and organizations. We believe that the four themes, under which relevant principles and norms can be

grouped, reflect what authors, organizations and working groups consider aspects of importance to governing data sharing activities in a responsible manner. These insights provide helpful leads for further work on conceptualising a harmonised governance framework for data sharing in health research. At the same time, our findings indicate that the current body of norms and principles is still too haphazard, non-uniform and sometimes even contradictory to serve as sufficient guidance in itself. Key questions, in particular how to deal with the limits of anonymity and how to effectuate meaningful public and patient involvement, will have to be part of the research agenda.

Appendix 1. Search queries (performed August 25, 2017).

Database	Search string	Returned
PubMed	(((((((international*[Title/Abstract]) OR supranational*[Title/Abstract]) OR ethic*[Title/Abstract]) OR moral*[Title/Abstract]) OR normative[Title/Abstract]) OR legal*[Title/Abstract])) AND ((((((((((Guideline*[Title/Abstract]) OR Guidance[Title/Abstract]) OR Code*[Title/Abstract]) OR Recommendation*[Title/Abstract]) OR Governance[Title/Abstract]) OR Declaration[Title/Abstract]) OR Regulatory[Title/Abstract]) OR Regulation*[Title/Abstract]) OR Framework[Title/Abstract])) AND ((((“big data”[Title/Abstract]) OR “data-sharing”[Title/Abstract]) OR “data-linkage”[Title/Abstract]) OR “data-intensive”[Title/Abstract]))	387
Embase	(‘data sharing’:ab,ti OR ‘big data’:ab,ti OR ‘data linkage’:ab,ti OR ‘data intensive’:ab,ti) AND (‘guideline*’:ab,ti OR ‘guidance’:ab,ti OR ‘code*’:ab,ti OR ‘recommendation*’:ab,ti OR ‘governance’:ab,ti OR ‘declaration’:ab,ti OR ‘regulatory’:ab,ti OR ‘regulation*’:ab,ti OR ‘framework*’:ab,ti) AND (‘international*’:ab,ti OR ‘supranational*’:ab,ti OR ‘normative’:ab,ti OR ‘moral*’:ab,ti OR ‘ethic*’:ab,ti OR ‘legal*’:ab,ti)	444
Google Scholar	allintitle: health OR framework OR governance OR recommendations “data sharing” [limit 2007-2017]	373
Scopus	TITLE-ABS (“data sharing” OR “data linkage” OR “data intensive” OR “big data”) AND TITLE-ABS (guideline* OR guidance OR code* OR recommendation* OR governance OR declaration OR regulation* OR regulatory OR framework*) AND TITLE-ABS (international* OR supranational OR normative OR moral* OR ethic* OR legal*) AND TITLE-ABS (medical OR health) AND TITLE-ABS (research)	215

Appendix 2. Breakdown of search terms.

"Data-sharing"	**AND**	**Guideline***	**AND**	**International***
"Data-linkage"		Guidance		Supranational*
"Data-intensive"		Code*		Normative
"Big Data"		Recommendation*		Moral*
		Governance		Ethic*
		Declaration		Legal*
		Regulatory		
		Regulation*		
		Framework		

Table 1. Selected ethical guidelines and recommendations.

Name of source	Organisation, Year	Status	Scope, Addressed to
International Ethical Guidelines for Health-related Research Involving Humans	Council for International Organizations of Medical Sciences (CIOMS), 2016	Ethical guideline, applies to activities designed to develop or contribute to generalizable health knowledge.	Universal scope, not defined who it is addressed to
Declaration of Taipei on Ethical Considerations Regarding Health Databases and Biobanks	World Medical Association (WMA), 2016	Ethical guideline, applies to the collection, storage and use of identifiable data and biological material beyond the individual care of patients.	Universal scope, primarily addressed to physicians. The WMA encourages others to adopt the principles.
Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects	World Medical Association (WMA), 2013	Ethical guideline, applies to medical research involving human subjects, including research on identifiable human material and data.	Universal scope, primarily addressed to physicians. The WMA encourages others to adopt the principles.
Framework for Responsible Sharing of Genomic and Health--Related Data	Global Alliance for Genomics and Health (GA4GH), 2014	A principled and practical framework, applies to the sharing of genomic and health-related data (for biomedical research)	Universal scope, addressed to all entities or individuals using genomic and health-related data.
The collection, linking and use of data in biomedical research and health care: ethical issues	The Nuffield Council on Bioethics, 2015	Report that sets out ethical principles and recommendations, related to the design and governance of data initiatives and data use for biological and medical research	United Kingdom / universal, addressed to anyone approaching a data initiative
Joint statement of purpose—vision, principles, and goals	Funders of public health research, 2011	Joint statement of funders, applies to sharing research data to improve public health	Universal, addressed to funders and the research community

Table 1. Continued.

Name of source	Organisation, Year	Status	Scope, Addressed to
Principles and Guidelines for Access to Research Data from Public Funding	Organisation for Economic Co-operation and Development (OECD), 2007	A legally non-binding recommendation, often referred to as soft law. Applies to research data that are gathered using public funds for the purposes of producing publicly accessible knowledge	Primarily addressed to OECD Member Countries and intended to assist all actors involved when trying to improve the international sharing of, and access to, research data
Recommendation of the Council on Human Biobanks and Genetic Research Databases	Organisation for Economic Co-operation and Development (OECD), 2009	A legally non-binding recommendation, often referred to as soft law. Provides guidance for the establishment, governance, management, operation, access, use and discontinuation of human biobanks and genetic research databases	OECD Countries, to be applied as broadly as possible, in particular to aid policy makers and practitioners who are establishing new human biobanks and genetic research databases. Can also usefully be applied to existing biobanks and databases
Recommendation of the Council on Health Data Governance	Organisation for Economic Co-operation and Development (OECD), 2017	A legally non-binding recommendation, often referred to as soft law. Applies to the access to, and the processing of, personal health data for health-related public interest purposes	OECD member countries and all levels of government, encourages non-governmental organisations to follow this Recommendation
Principles for Responsible Clinical Trial Data Sharing: Our Commitment to Patients and Researchers	The European Federation of Pharmaceutical Industries and Associations (EFPIA) and the Pharmaceutical Research and Manufacturers of America (PhRMA), 2013	Joint policy of organizations representing pharmaceutical industries	Members of EFPIA and PhRMA

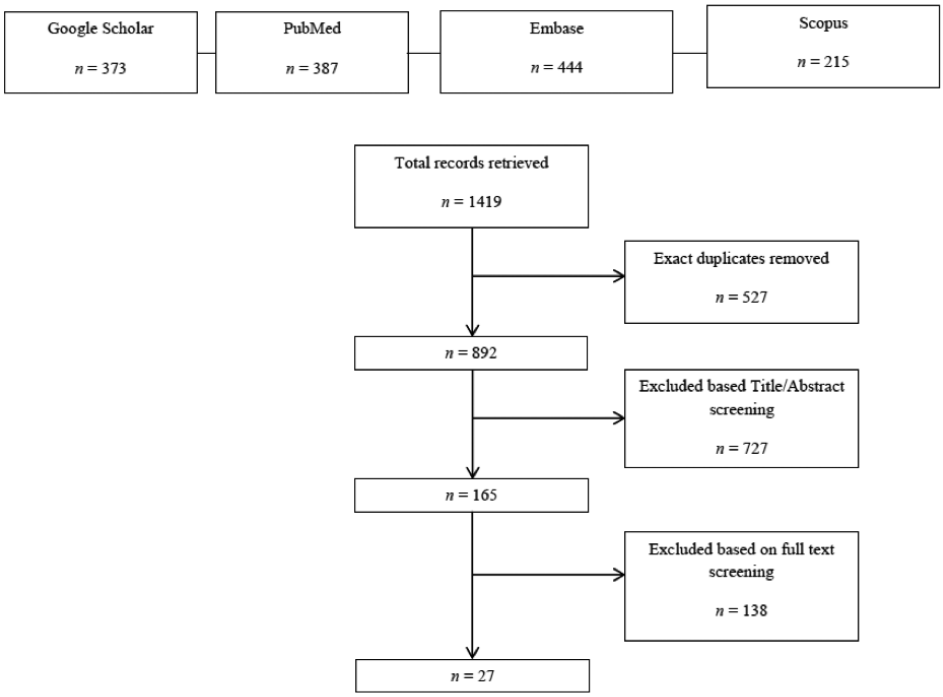

Figure 1 Flow diagram of the selection and inclusion of publications.

Table 2. Themes and principles

Main themes	Norms and principles
Societal benefits and value	Data quality[7,12,15–17,20,21,25,26,29] Accessibility[6,7,12,16,20,22,26,28,30] Sustainability[7,15,16,20,22,25,26] Scientific progress/value[7,9,14,16] Promote health and well-being[7,23,32] Scientific validity[22,25,29] Collaboration and capacity building[14,19] Interoperability[15,16] Societal benefit[19,27] Health-related public interest[19] Social value[14] Improve public health[9] Efficiency[15] Duty to share[6]

Table 2. Continued.

Main themes	Norms and principles
Distribution of risks, benefits and burdens	Benefit sharing[16,20,25,26] Reciprocity[7,20,26,29] Risk-benefit evaluation[13,14,16,32] Equity[7,9,14] Proportionality[21,25,26] Intellectual property[15,29,30] Attribution[20,26,30] Ownership[18,33] Recognition and attribution[7] Individual benefit[19]
Respect for individuals and groups	Respect/protect privacy[8,9,12,16,19–22,25–29,32,33,35,36,37,39] Protect confidentiality[6,8,16,18,20,25–27,32,37] Data security[7,15,16,18,20,21,26,34,39] Respect individuals[7,13,27,32] Risk-benefit evaluation[13,14,16,32] Respect individual rights[13,14,16,32] Individual autonomy[8,27,29,32] Respect (the dignity of) communities[7,9] Respect dignity of individuals[27,32] Legal compliance[21,22] Prevent discrimination[16,27] Protect life, health and well-being[32] Respect families[7] Respect welfare of individuals[14]
Public trust and engagement	Accountability[7,12–16,18–22,25–28] Engagement / participation[7,12–14,16,19–22,27,28] Transparency[7,12,15,16,19,20,25–27,38,39] Maintain public trust[12,19–22,27,28] Responsibility[18,20,22,25,34,37] Maintain integrity[12,20,21,26,27] Professionalism[7,15,16,19] Health democracy[25] Solidarity[41]

Table 3. Informing and enabling participants and the public

Potential participants need to be informed about:
the limits to anonymity and confidentiality of data;[7,13,14,16]
the type of research being carried out, the activities of health databases and/or the research results;[7,16,27]
how consent can be withdrawn, as well as the implications of and limits to withdrawal;[14,16,27]
whether return of individual-level findings derived from analysis of the data is foreseen and the right to opt-out from receiving such information;[14,16]
how the data and the confidentiality of these data will be protected;[14,16]
the legal basis and objectives of the data processing by third parties;[19]
whether the participants retain any rights over the data;[16]
the exceptional circumstances and conditions under which researchers may access data that is not coded or anonymous;[16]
the potential adverse consequences of breaches of confidentiality;[14]
information about an actual significant data breach or misuse of data;[19]
significant modifications to databases' policies, protocols and procedures;[16]
entering into commercial collaborations or commercialisation of research resources.[16]
Enable participants to exercise the following rights:
the right to withdraw consent;[14,16,19,27]
the right to choose whether (and how) individual-level findings will be returned;[14,16]
the right to request for information about their data and its use;[27]
the right to request for corrections of omissions in data;[27]
the choice to opt-out of being re-contacted for research purposes.[14]
Related to data sharing, public information should include the following items:
the terms, procedures, policies and/or governance frameworks for data access or sharing;[7,13,15,16,19]
for what purposes and ways in which data may be shared;[7,13,19]
a summary of (approved) data transfers,[7] including a list of categories of approved data recipients;[19]
the legal bases for sharing data;[19]
a catalogue of the resources accessible for research purposes;[16]
the duration of data storage;[7]
a specification of conditions attached to the use of the data;[15]
a summary of research results;[16]
commercial involvement and propriety claims;[7]
processes of withdrawal from data sharing;[7]
contact information and answers to frequently asked questions;[16]
procedures for handling complaints;[27]
the purpose, background, funding, scope, uncertainties and risks, scientific rationale of the initiative or database and its funding;[16]
the disclosure of any conflict of interest involving personnel.[16]

Table 4. Items subject to ethical review

A REC (or a comparable ethical review body) should review:
- (the justification of) a waiver of informed consent requirements;[14,16,27,32]
- whether the consent given is sufficient for the planned use;[14,27]
- for determining when to seek re-consent;[16]
- use of data on the basis of broad consent;[16]
- Usage of data not anticipated in the original informed consent process;[16]
- Re-use in cases where informed consent may not have been obtained previously;[16]
- whether the consent procedure meets the specifications of broad informed consent;[14]
- whether explicit informed consent is required;[14]
- whether an informed opt-out procedure can be used;[14]
- the proposed usage and/or collections, the storage protocol;[14]
- if other measures need to be taken to protect the donor;[27]
- the use of personal identifiers, its necessity and how confidentiality will be protected;[14]
- whether individual counselling is necessary when returning genetic findings.[14]

References

1. Hemingway H, Asselbergs FW, Danesh J et al. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. Eur Heart J 2017. doi:10.1093/eurheartj/ehx487.
2. BD4BO – Big Data for Better Outcomes. Available at http://bd4bo.eu/.
3. Dove ES, Thompson B, Knoppers BM. A step forward for data protection and biomedical research. Lancet 2016; 387: 1374–75.
4. Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care.data ran into trouble. J Med Ethics 2015; 41: 404–9.
5. Laurie G. Reflexive governance in biobanking: on the value of policy led approaches and the need to recognise the limits of law. Hum Genet 2011; 130: 347–56.
6. Alfonso F. Data sharing: A new editorial initiative of the International Committee of Medical Journal Editors. Netherlands Hear J 2017; 25: 297–303.
7. Global Alliance for Genomics and Health (GA4GH). Framework for Responsible Sharing of Genomic and Health Related Data. 2014. Available at https://www.ga4gh.org/ga4ghtoolkit/regulatoryandethics/framework-for-responsible-sharing-genomic-and-health-related-data/
8. Dyke SO, Hubbard TJ. Developing and implementing an institute-wide data sharing policy. Genome Med 2011; 3: 60.
9. Funders of public health research. Joint statement of purpose—vision, principles, and goals. 2011. Available at https://wellcome.ac.uk/what-we-do/our-work/sharing-research-data-improve-public-health-full-joint-statement-funders-health
10. Auffray C, Balling R, Barroso I et al. Making sense of big data in health research: Towards an EU action plan. Genome Med 2016; 8: 71.
11. Lea NC, Nicholls J, Dobbs C et al. Data Safe Havens and Trust: Toward a Common Understanding of Trusted Research Platforms for Governing Secure and Ethical Health Research. JMIR Med informatics 2016; 4: e22.
12. Baker DB, Kaye J, Terry SF. Privacy, Fairness, and Respect for Individuals. eGEMs (Generating Evid Methods to Improv patient outcomes) 2016; 4: 7.
13. The Nuffield Council on Bioethics. The collection, linking and use of data in biomedical research and health care: ethical issues. 2015. Available at http://nuffieldbioethics.org/project/biological-health-data
14. Council for International Organizations of Medical Sciences (CIOMS). International Ethical Guidelines for Health-related Research Involving Humans. 2016. Available at https://cioms.ch/shop/product/international-ethical-guidelines-for-health-related-research-involving-humans/

15. Organisation for Economic Co-operation and Development (OECD). Principles and Guidelines for Access to Research Data from Public Funding. 2007. Available at https://www.oecd-ilibrary.org/science-and-technology/oecd-principles-and-guidelines-for-access-to-research-data-from-public-funding_9789264034020-en-fr
16. Organisation for Economic Co-operation and Development (OECD). Recommendation of the Council on Human Biobanks and Genetic Research Databases. 2009. Available at http://www.oecd.org/sti/emerging-tech/guidelines-for-human-biobanks-and-genetic-research-databases.htm
17. Rodriguez H, Snyder M, Uhlén M et al. Recommendations from the 2008 International Summit on Proteomics Data Release and Sharing Policy: The Amsterdam Principles. J Proteome Res 2009; 8: 3689–92.
18. Chan T, Di Iorio CT, De Lusignan S, Lo Russo D, Kuziemsky C, Liaw S-T. UK National Data Guardian for Health and Care's Review of Data Security: Trust, better security and opt-outs. J Innov Heal Informatics 2016; 23: 627.
19. Organisation for Economic Co-operation and Development (OECD). Recommendation of the Council on Health Data Governance. 2017. Available at http://www.oecd.org/els/health-systems/health-data-governance.htm
20. Knoppers BM. Framework for responsible sharing of genomic and health-related data. Hugo J 2014; 8: 3.
21. Dove ES, Knoppers BM, Zawati MH. An ethics safe harbor for international genomics research? Genome Med 2013; 5: 99.
22. Antman EM, Benjamin EJ, Harrington RA et al. Acquisition, Analysis, and Sharing of Data in 2015 and Beyond: A Survey of the Landscape. J Am Heart Assoc 2015; 4: e002810.
23. EFPIA, PhRMA. Principles for Responsible Clinical Trial Data Sharing: Our Commitment to Patients and Researchers. 2013.Available at http://phrma-docs.phrma.org/sites/default/files/pdf/PhRMAPrinciplesForResponsibleClinicalTrialDataSharing.pdf.
24. Allen C, Des Jardins TR, Heider A et al. Data governance and data sharing agreements for community-wide health information exchange: lessons from the beacon communities. EGEMS 2014; 2: 1057.
25. Bredenoord AL, Mostert M, Isasi R, Knoppers BM. Data sharing in stem cell translational science: policy statement by the International Stem Cell Forum Ethics Working Party. Regen Med 2015; 10: 857–61.
26. Regulatory and Ethics Working Group, Global Alliance for Genomics & Health R and EW, Sugano S, Sugano S. International code of conduct for genomic and health-related data sharing. Hugo J 2014; 8: 1.
27. World Medical Association (WMA). Declaration of Taipei on Ethical Considerations Regarding Health Databases and Biobanks. 2016. Available at https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks/
28. Laurie G, Sethi N. Towards Principles-Based Approaches to Governance of Health-related Research using Personal Data. Eur J risk Regul EJRR 2013; 4: 43–57.

29. Mascalzoni D, Dove ES, Rubinstein Y et al. International Charter of principles for sharing bio-specimens and data. Eur J Hum Genet 2015; 23: 721–8.
30. Chokshi DA, Parker M, Kwiatkowski DP. Data sharing and intellectual property in a genomic epidemiology network: policies for large-scale research collaboration. Bull World Health Organ 2006; 84: 382–7.
31. Duchange N, Darquy S, d'Audiffret D et al. Ethical management in the constitution of a European database for leukodystrophies rare diseases. Eur J Paediatr Neurol 2014; 18: 597–603.
32. World Medical Association (WMA). Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects. 2013. Available at https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/
33. Andrew NE, Sundararajan V, Thrift AG et al. Addressing the challenges of cross-jurisdictional data linkage between a national clinical quality registry and government-held health data. Aust N Z J Public Health 2016; 40: 436–42.
34. Shenkin SD, Pernet C, Nichols TE et al. Improving data availability for brain image biobanking in healthy subjects: Practice-based suggestions from an international multidisciplinary working group. Neuroimage 2017; 153: 399–409.
35. Dyke SO, Dove ES, Knoppers BM. Sharing health-related data: a privacy test? Genomic Med 2016; 1: 16024.
36. Banzi R, Bertele' V, Demotes-Mainard J et al. Fostering EMA's transparency policy. Eur J Intern Med 2014; 25: 681–4.
37. Tucker K, Branson J, Dilleen M et al. Protecting patient privacy when sharing patient-level data from clinical trials. BMC Med Res Methodol 2016; 16: 77.
38. Kostkova P, Brewer H, de Lusignan S et al. Who Owns the Data? Open Data for Healthcare. Front public Heal 2016; 4: 7.
39. ACMG Board of Directors AB of. Laboratory and clinical genomic data sharing is crucial to improving genetic health care: a position statement of the American College of Medical Genetics and Genomics. Genet Med 2017; 19: 721–2.
40. Sharer JD, Bodamer O, Longo N, Tortorelli S, Wamelink MMC, Young S. Laboratory diagnosis of creatine deficiency syndromes: a technical standard and guideline of the American College of Medical Genetics and Genomics. Genet Med 2017; 19: 256–63.
41. Prainsack B, Buyx A. A solidarity-based approach to the governance of research biobanks. Med Law Rev 2013; 21: 71–91.
42. Paltoo DN, Rodriguez LL, Feolo M et al. Data use under the NIH GWAS Data Sharing Policy and future directions. Nat Genet 2014; 46: 934–8.
43. Childress JF, Faden RR, Gaare RD et al. Public Health Ethics: Mapping the Terrain. J Law, Med Ethics 2002; 30: 170–8.
44. Global Alliance for Genomics and Health (GA4GH). Regulatory & Ethics Toolkit. Available at https://www.ga4gh.org/ga4ghtoolkit/regulatoryandethics/.

45. BBMRI-ERIC. A Code of Conduct for Health Research. Available at http://code-of-conduct-for-health-research.eu/.
46. Childress JF, Faden RR, Gaare RD, Gostin LO, Kahn J, Bonnie RJ, Kass NE, Mastroianni AC, Moreno JD, Nieburg P. Public Health Ethics: Mapping the Terrain. J Law, Med Ethics Wiley/Blackwell (10.1111); 2002;30:170–178.
47. GSK. Data transparency | GSK. 2014. Available at https://www.gsk.com/en-gb/behind-the-science/innovation/data-transparency/.
48. European Medicines Agency (EMA). European Medicines Agency policy on publication of clinical data for medicinal products for human use. 2014. Available at http://www.ema.europa.eu/docs/en_GB/document_library/Other/2014/10/WC500174796.pdf.

Chapter 7

Data sharing in stem cell translational science: Policy statement by the International Stem Cell Forum Ethics Working Party

Bredenoord AL
Mostert M
Isasi R
Knoppers BM

Regenerative Medicine 2015; 10: 857-61.

Abstract

Data and sample sharing constitute a scientific and ethical imperative but need to be conducted in a responsible manner in order to protect individual interests as well as maintain public trust. In 2014, the Global Alliance for Genomics and Health (GA4GH) adopted a common Framework for Responsible Sharing of Genomic and Health-Related Data. The GA4GH Framework is applicable to data sharing in the stem cell field, however, interpretation is required so as to provide guidance for this specific context. In this paper, the International Stem Cell Forum Ethics Working Party discusses those principles that are specific to translational stem cell science, including engagement, data quality and safety, privacy, security and confidentiality, risk–benefit analysis and sustainability.

Background

In 2010, the Stem Cell Charter affirmed five Principles as the foundation of responsible stem cell research:[1]

- Responsibility to maintain the highest level of scientific quality, safety and ethical probity;
- Protection of citizens from harm and the safeguarding of public trust and values;
- Intellectual freedom to exchange ideas in the spirit of international collaboration;
- Transparency through the disclosure of results and of possible conflicts of interest;
- Integrity in the promotion and advancement of stem cell research and therapy for the betterment of the welfare of all human beings.

Since 2010, the stem cell field has thrived, particularly due to the convergence of stem cell science with genomic technologies enabling intense genetic characterization of stem cell lines. Today, the sharing of genomic- and health-related data for biomedical research to achieve statistical significance and to foster translational medicine in the field of stem cell science is of the utmost importance. Indeed, international data sharing stimulates scientific progress and is more efficient and economical. Yet, in the absence of proper governance and security, the risk of privacy infringements of research participants and their family members may increase. Data and sample sharing constitutes a scientific and ethical imperative but needs to be conducted

in a responsible manner in order to protect individual interests as well as maintain public trust.[2]

In 2014, the Global Alliance for Genomics and Health (GA4GH) adopted a common Framework for Responsible Sharing of Genomic and Health-Related Data (hereinafter the 'Framework'). The goal is to develop harmonized approaches both to enable effective and responsible sharing of genomic and clinical data and to catalyze data sharing projects that drive and demonstrate its value. A prerequisite for applying this Framework in practice is that the use of data is in compliance with national and international laws, general ethical principles and best practice standards that respect restrictions on downstream uses. In the Framework, the GA4GH adopted four Foundational Principles for responsible data sharing:

- Respect individuals, families and communities;
- Advance research and scientific knowledge;
- Promote health, wellbeing and the fair distribution of benefits;
- Foster trust, integrity and reciprocity.

The GA4GH Framework is applicable to data sharing in the stem cell field, however, adaptation and interpretation are required so as to provide guidance for this specific context. Here, we apply the principles and core elements of the GA4GH Framework to the context of stem cell science. Although all the (interconnected) principles of the Framework and Stem Cell Charter apply, the International Stem Cell Forum (ISCF) Ethics Working Party here limits itself to further discussion of those principles that are specific to the aims of translational stem cell research.

Context: translational stem cell science

Although stem cells come in all shapes and sizes, they generally are characterized by two properties: their capacity for both self-renewal and for differentiation into specialized cell types. Stem cells can originate from embryonic, fetal or adult tissue and are broadly categorized accordingly. Pluripotent stem cells (PSCs) are capable of self-renewal and have the capacity to differentiate into any cell type of the human body. In 1998, scientists first isolated human embryonic stem cells (hESCs) from the inner cell mass of the early human embryo.[3] Another way to generate PSCs was discovered in 2006, when researchers showed that differentiated fibroblasts could be reprogrammed into stem cells capable of forming all three germ layers.[4,5]

Translational stem cell science refers to research that focuses on the biology and application of all types of stem cells. It is widely perceived as offering promising therapeutic potential to develop innovative treatment for a variety of disorders, ranging from neurodegenerative conditions to cardiovascular disease and cancer.[6] The search for and development of stem cell based therapies forms one of the cornerstones of the novel field of regenerative medicine as it is focused on repair, replacement, or regeneration of cells, tissues or organs to restore impaired function. Regenerative medicine is characterized by its interdisciplinary nature: stem cell biology, genetics, material sciences, bioinformatics and surgery work together in this field. Specific characteristics of regenerative medicine such as the potential immortality of cell lines, the complexity of the interventions, the new aim of regeneration, the tremendous scientific and commercial stakes and the high public attention give a new twist to the classical challenges of research ethics.[7]

The clinical translation of stem cells is scientifically and ethically challenging.[7] Registries and repositories are vital infrastructures for both basic stem cell science and the translation toward regenerative medicine. Generally, they include primary material such as human tissue samples, cell lines and associated data. Robust banking networks and registries enable global access to well-characterized and traceable PSC lines.[2] The accessibility of data and information associated with such cell lines, including provenance documentation, technical information and intellectual property rights, are critical to their management and utility.[8,9]

Specific policy principles

Engagement

- Stem cell professionals are encouraged to participate in the debate on the ethical and societal implications of the use and sharing of data to further regenerative medicine. Moreover, as science, society and technology are mutually constitutive, the active involvement, training and education of stem cell scientists will co-shape the societal impact of regenerative medicine and drive responsible innovation;[7]
- Active involvement of citizens in the public understanding of stem cell research and of the need for data sharing serves to increase public understanding and potentially, to democratize the direction of such research and its oversight;[10]
- A participatory approach to research should be encouraged, in which donors and patients are involved in (clinical) research from the beginning throughout clinical trials. This is in line with the recognition that stem cell research is a

social endeavour. Engagement could also reduce the therapeutic misconception, which is particularly present in stem cell based interventions where involvement in research is often perceived as a medical 'treatment';
- Ongoing debate regarding the wider and future societal implications, limits and impacts of the use of data in stem cell research and regenerative medicine should be encouraged.

Data quality & safety
- Maintaining the highest level of quality assurance and safety are of the utmost importance in translational stem cell science.[1] Provenance determination of cell lines also ensures ethical probity, transparency and accountability;
- General information about the nature of stem cell line derivation and use by researchers should be publicly documented and available, ideally, in international registries;
- All proposed protocols for data sharing should aim to generate new knowledge and understanding using rigorous scientific methods. Research findings, including negative results, should be published in order to allow reproducibility and further research;
- Data element standards should be promoted to enable the comparison of shared data.[11]

Privacy, security & confidentiality
- The implications of data sharing on privacy and confidentiality should be explicitly addressed in the consent process, while acknowledging the impossibility of guaranteeing absolute privacy.[2] Stem cell science in the era of increasing data-intensive medical research requires continuous attention for novel approaches to protect privacy, for example, security technologies and de-identification mechanisms. Attention should be given to novel safeguards for protecting sensitive data. Prospectively adopting protocols for providing general (aggregate) research results is encouraged.[12] A policy for returning individual results or findings (when appropriate and desired) for PSC lines is also recommended. A key component of this policy is a protocol for disclosure, approved by an independent ethics review committee or by an oversight committee. The protocol should describe the mechanisms and conditions for appropriate disclosure, including the scope of the responsibilities of all the stakeholders involved. The individual

should have consented to such disclosure, the results or findings be validated, have clinical utility and be actionable;[10]

- The ISCF Ethics Working Party cautions against any return of donor-specific results to embryo and gamete donors, given the nature of human embryonic stem cell lines and the circumstances of their derivation.[12]

Risk–benefit analysis

- Proportionality in risk–benefit analysis is crucial, due not only to the perceived and actual benefits and risks of stem cell translational research, but also due to the vulnerability of patients and the scientific credibility of the field itself;
- A proportional assessment of the risks of individual identifiability should be tailored to the nature of cell line derivation (e.g., hESCs vs iPSCs). Such an assessment requires an evaluation of real risks and potential benefits and social value.[2,13]
- A cautious approach is needed when sharing raw sequence reads (such as whole genomes) given that they contain personal information that is directly identifiable or would facilitate re-identification.[2]

Sustainability

- There is a need to underscore the importance of the long-term sustainability of scientific infrastructures, such as stem cell banks and registries. This will ensure not only their maintenance but also to respect the wishes of research participants and patients concerning the use(s) of their data and samples;
- Oversight mechanisms should be in place for ongoing robust governance. Attention should be paid to prospectively address the issues specific to stem cell research;
- Novel types of consent have been proposed for research with biological samples and cell lines, one of these being broad consent. We understand broad consent to future unspecified research (subject to ethics approval) as consent for governance. This means that broad consent is specifically aimed at providing the donor with information on the governance structure of the biobank in question.[14] This should for example include information regarding ethical oversight, property rights and potential commercial use, and also include information on how the data and samples are stored, accessed and used.

Sunset clause

The ISCF will revisit this policy every 3 years. It is not our aim to make standards, but rather to propose best practices for this area. We encourage researchers to integrate this policy into their protocols and professional guidelines.

Future perspective

The sharing of genomic and health-related data for biomedical research to achieve statistical significance and to foster translational medicine in the field of stem cell science will be of increasing importance. Yet, in the absence of proper governance and security, the risk of privacy infringements of research participants and their family members may increase. Data and sample sharing constitute a scientific and ethical imperative but need to be conducted in a responsible manner in order to protect individual interests as well as maintain public trust. In 2014, the Global Alliance for Genomics and Health (GA4GH) adopted a common Framework for Responsible Sharing of Genomic and Health-Related Data. We applied the GA4GH Framework to translational stem cell science. We encourage researchers to integrate this policy into their protocols and professional guidelines in order to stimulate responsible innovation in translational stem cell science.

References

1. Knoppers BM, Isasi R, Willemse L. Stem cell charter. Regen. Med. 2010; 5: 5–6.
2. Isasi R, Andrews PW, Baltz JM et al. Identifiability and privacy in pluripotent stem cell research. Cell Stem Cell 2014; 14: 427–30.
3. Thomson JA, Itskovitz-Eldor J, Shapiro SS et al. Embryonic stem cell lines derived from human blastocysts. Science 1998; 282: 1145–7.
4. Takahashi K, Yamanake S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 2006; 126: 663–76.
5. Takahashi K, Tanabe K, Ohnuki M et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 2007; 131: 861–72.
6. Gögel S, Gubernator M, Minger SL. Progress and prospects: stem cells and neurological disease. Gene Ther. 2010; 18: 1–6.
7. Niemansburg SL, Teraa M, Hesam H, Van Delden JJM, Verhaar MC, Bredenoord AL. Stem cell trials for vascular medicine: ethical rationale. Tissue Eng. Part A 2014; 20: 2567–74.
8. The Hinxton Group. Statement on Policies and Practices Governing Data and Materials Sharing and Intellectual Property in Stem Cell Science. Available at www.hinxtongroup.org/consensus_hg10_final.pdf
9. Matthews DJH, Graff GD, Saha K, Winickoff DE. Access to stem cells and data: persons, property rights, and scientific progress. Science 2011; 331: 725–7.
10. Bredenoord AL, Onland-Moret NC, Van Delden JJM. Feedback of individual genetic results to research participants: in favor of a qualified disclosure policy. Hum. Mutat. 2011; 32: 1–7.
11. Andrews PW, Baker D, Benvinisty N et al. Points to consider in the development of seed stocks for pluripotent stem cells for clinical application: International Stem Cell Banking Initiative. Regen. Med. 2015; 10(Suppl. 2): S1–S44.
12. Isasi R, Knoppers BM, Andrews PW et al. Disclosure and management of research findings in stem cell research and banking: policy statement. Regen. Med. 2012; 7: 439–48.
13. Habets MGJL, Van Delden JJM, Bredenoord AL. The social value of clinical research. BMC Medical Ethics 2014; 15: 66.
14. Boers SN, Van Delden JJM, Bredenoord AL. Broad consent is consent for governance. Am. J. Bioethics 2015; 15; 53–5.

Chapter 8

General discussion

Introduction

Big Data is transforming health research into a data-intensive endeavour. This has sparked a lively discussion about how such data use should be governed by principles and rules.[1] Moreover, the need for enhanced protection of (personal) data has led to a profound legislative reform of data protection law in the European Union (EU), which resulted in the adoption of the much debated General Data protection Regulation (GDPR).[2,3]

The main aim of this thesis is to inform the debate about what form laws, regulations and information governance should take in the EU, to allow for progress in data-intensive health research while safeguarding (fundamental) rights and morally relevant interests. To achieve this aim, this thesis addresses the central question of how relevant rights and interests can be safeguarded and balanced in the EU, without disproportionately hampering data-intensive health research.

Main conclusions

In chapters 2, 3 and 4, the key challenges and ways forward in the EU legal framework on privacy and data protection, which are relevant to data-intensive health research, have been discussed. Subsequently, chapters 5, 6 and 7 provide insight in other relevant sources of normativity, and show how concerns can be dealt with in specific contexts.

Following a recent change in the EU legal framework on the fundamental rights level, related to privacy and data protection, chapter 2 addresses the question: '*Are there differences between the right to data protection and the right to privacy in the EU, which are relevant in the context of data-intensive health research?*' It is shown that there is indeed relevance in taking the differences between the right to privacy and the right to data protection into account, also in the context of data-intensive health research. It aims to complement the right to privacy, by positively guaranteeing a more comprehensive and harmonised system of data protection norms. Such a comprehensive system of data protection should be considered to serve two functions in particular. Firstly, the aim is to provide effective overarching safeguards that secure the rights and interests of individuals, irrespective of whether the personal data processing is grounded on consent or any other legal basis. The overarching safeguards should, amongst other things, include requirements of accountability subject to independent oversight, transparency towards data subjects and the public,

ensure that data subjects can invoke their rights, and data security. Secondly, the system of data protection arranges for specific exemptions and safeguards related to data processing in scientific research. These specific safeguards should compensate for the loss of individual control as a result of the research exemptions from consent requirements, certain general principles, and individual rights.

During the reform of EU data protection law, there was a lively debate about informed consent and anonymity. To contribute to this debate, chapter 3 addresses the question: '*How is the consent or anonymise approach challenged in a data-intensive health research context, and what are possible ways forward within the EU legal framework on data protection?*' The 'consent or anonymise approach' has been dominant in many health research initiatives.[3,4] According to this approach, only two options are available when personal health data are to be shared or accessed for research purposes. The first option is that the data are processed on the basis of informed consent. The second option is that extensive de-identification measures are taken so that the shared data can no longer be considered personal data. Potential solutions to data sharing or access problems can therefore, according to the consent or anonymise approach, only be found by modifying or tweaking concepts of informed consent or anonymity. Although there is value in thinking about how the concept or process of informed consent or anonymisation could be improved, and promising ways forward are being explored,[5,6] it should be prevented that other legal bases than informed consent are disregarded beforehand. Moreover, an overly strong focus on requirements of consent or anonymity could distract the attention from the debate on and implementation of other essential measures to protect, promote and balance relevant rights and interests. One of these essential measures on the national level is the implementation of a research exemption from consent in data protection law. A research practice based on consent exemptions, accompanied with appropriate safeguards, should be preferred above continuing a practice of stretching concepts of consent or anonymisation beyond their limits.

After years of uncertainty about how the GDPR would impact data-intensive health research, the EU legislative bodies adopted the final version of the GDPR. This sparked the question: '*Does the GDPR contribute to a responsible and effective use of personal data in data-intensive health research?*' in chapter 4. After a thorough evaluation of the final version of the GDPR, I conclude that overall the GDPR does contribute to a responsible and effective use of personal data in data-intensive health research.[7] I argue that the GDPR implements a system of data protection norms that safeguards the use of personal data in health research without disproportionately

hampering such usage. The many open norms in the GDPR sufficiently allow for a data protection regime that is specifically tailored to the context of health research. Moreover, the large number of research exemptions and derogations combined with a specific regime of conditions and safeguards could prevent that research is disproportionately hampered. Nevertheless, it should also be noted that the many open norms, and the research exemptions that need to be implemented in national law, do not contribute to a high level of harmonisation. This could negatively impact a coherent implementation and interpretation of the GDPR on the national level, and therefore potentially hinder international data sharing and access. Such a negative impact could partially be mitigated by formally recognised codes of conduct conform Article 40 GDPR, like the Code of Conduct for Health Research that is being developed under the auspices of BBMRI-ERIC.[8]

To gain more insight into how concerns related to privacy and data protection are dealt with in practice, chapter 5 addresses the question: *What challenges related to privacy- and data protection are encountered in real-world examples of data-intensive health research?* We conducted a multiple-case study in which the YOUth cohort, a longitudinal cohort focusing on psychosocial development, and Big Data Psychiatry, a pilot study in Big Data analytics on psychiatric health data, were selected as cases. Three themes emerged from the analysis. The first theme concerns the move away from anonymisation as a strategy to prevent the applicability of data protection law. Alternative measures have been implemented with the aim of complying with the law and mitigating risks. The second theme relates to the search for meaningful and proportionate ways to allow individuals to control and be aware of the use of their data. This search explores the possible use of digital tools to ensure a more ongoing engagement with participants. Thirdly, uncertainty about how to comply with the law is perceived as having negative impact on the initiatives. This uncertainty is primarily attributed to a lack of easily accessible guidance and to the recent changes in data protection law. Overall, the findings show that it takes considerable effort to take privacy and data protection norms into account in practice, especially when individual level data need to be linked or enriched. What is more, researchers often struggle with uncertainties and legal complexity. By mapping the complex interplay between legal norms and practical requirements in an early phase, setbacks could be prevented, the flexibility within the law could be found, and systems and organisations could be designed while taking relevant norms into account.

Data-intensive health research is not merely governed by legal norms. Complementary to the legal analysis, chapter 6 systematically reviews guidelines

and academic literature to answer the question: *What are the ethically relevant principles and norms so far developed by (international) working groups or professional organisations with respect to data sharing in health research?* We observed a broad range of principles and norms with considerable convergence at the aggregate level of four overarching themes: societal benefits and value; distribution of risks, benefits and burdens; respect for individuals and groups; and public trust and engagement. Though providing helpful leads for a governance framework for international data sharing, the current guidance is often too varied, haphazard and sometimes even contradictory. An example is that anonymisation is discouraged in one of the sources, while other sources prefer or demand complete de-identification. Another example is the diversity of and contradiction between norms related to how and whether informed consent should be obtained. Our work highlights the need for considerable investments and expertise to further develop and implement a governance framework for international data sharing initiatives. More specifically, there is scarce guidance on how to deal with the limitations to preserve anonymity and confidentiality of personal data. A common recommendation is to inform participants and the public about these limitations. Future work should concentrate more on measures to safeguards rights and interests and secure public trust, at all levels of (de-) identification.

In chapter 7, stem cell science serves as an example of how principles and norms need to be transposed into policy a context-specific way. The following question is answered: *What are the specific policy principles for responsible data sharing in stem cell translational science?* In the field of today's stem cell science, the sharing of genomic- and health-related data for biomedical research is of the utmost importance. Data and sample sharing constitute a scientific and ethical imperative, but need to be conducted in a responsible manner. In 2014, the Global Alliance for Genomics and Health (GA4GH) adopted a common Framework for Responsible Sharing of Genomic and Health-Related Data. In our report consisting of a policy statement, this Framework is further interpreted and specified to provide guidance for the specific context of translational stem cell science.[9] Some of the key principles need a tailored interpretation, including engagement, data quality and safety, privacy, security and confidentiality, risk-benefit analysis and sustainability. We encourage researchers to integrate this policy in their guidelines and protocols, in order to stimulate responsible innovation.

Chapter 8

Perspectives on ways forward

Although we mapped and discussed many of the important normative issues in data-intensive health research, the search for appropriate ways forward remains an ongoing process. This is not only a consequence of a lack of agreement in the academic debate, but is also inherent to the complex and constantly evolving normative and research landscapes. Data protection law for instance, is often criticised for its complexity. Complexity should, however, not always be considered as a negative aspect of the law. On the contrary, a certain level of complexity is essential for the functioning of the law, also in the context of data-intensive health research.[10] Embracing legal complexity in practice could allow for a more flexible, proportionate and just interpretation and application of the law. Here, we discuss and recommend some key elements of such an approach, in which both legal and practical complexities are taken into account.

Flexibilities in the law

Especially in international collaborations, also within the EU, a failure to make use of the flexibilities in the law could result in an overly defensive approach to data access and sharing. Making use of flexibilities in the law should not be about finding loopholes to circumvent inconvenient or much demanding legal obligations. It is an essential element of the search for a proportionate way forward in data-intensive health research, with due consideration of the aims and central principles of the law. When it comes to the GDPR, these aims include both the protection of fundamental rights *and* advancing scientific knowledge in the public interest. The GDPR has not been introduced in the EU to prohibit or unduly hamper data access and sharing for scientific research purposes. On the contrary, the GDPR aims to facilitate scientific research and confirms that personal data can be processed for scientific research purposes, subject to appropriate conditions and safeguards.[11]

This special status of scientific research in the GDPR, and the related flexibility, is in particular reflected in the various research exemptions. The research exemptions include exemptions or derogations from some of the general principles, consent requirements and individual rights.[3,7] Making use of such derogations comes with the responsibility to meet certain conditions and implement a set of organisational and technical measures. Once these conditions are met and sufficient measures are taken, in particular by adhering to the principle of data minimisation, it is allowed to invoke the flexibility that these research exemptions provide. Another example

of flexibility within the GDPR is the interpretation of the requirement of informed consent in the context of scientific research. At first glance, it would seem that the requirement of 'specific' informed consent would not allow for 'broad' consent for a range of purposes. Recital 33 of the GDPR, however, aids in the interpretation of this article and explains that "*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*". Therefore, in the context of health research, it is allowed to obtain broad consent for a range of research purposes as long as recognised ethical standards are followed. This implies that recognised ethical standards such as those laid down in the CIOMS guidelines and the Declaration of Taipei gain legal relevance.[12,13] These guidelines emphasize that the ethical acceptability of broad informed consent relies on proper governance structure. Many of the (other) open norms in the GDPR, however, still need further explanation about how they should be interpreted or applied when personal data are processed for research purposes.[7] Nevertheless, some additional legal guidance and certainty could be provided by means of approved codes of conduct (Article 40 GDPR),[8] or by the supervisory authorities and European Data Protection Board (Chapters VI and VII GDPR).

Legal compliance and beyond

When the law is applied to personal data processing activities in a specific data-intensive health research initiative, questions about legal compliance arise. The importance of fostering and demonstrating legal compliance has increased substantially with the introduction of the GDPR. The accountability principle and rules in the GDPR clarify and strengthen the responsibility for complying with the GDPR, and require more to demonstrate compliance. What is more, substantial administrative fines could be imposed on those who fail to comply.[14] Adhering to the GDPR and implementing measures that foster legal compliance, which do not disproportionately hinder research, is a daunting task for several reasons. As indicated above, both the regulatory and the research landscape are complex and constantly evolving. What is more, there seems to be a lack of clear guidance and accessible expertise on how privacy and data protection issues should be dealt with.[10] It therefore takes considerable efforts to take data protection and privacy-related norms into account in practice. In addition, there is a danger that striving for legal compliance unnecessary hampers data sharing and access, especially when legal issues are dealt with in a reactive way.

In the search for an effective and proportionate approach to legal compliance, it is essential to proactively identify and address data protection related issues, so that data protection enhancing and preserving measures can be embedded in the design of systems and organisations. This form of value-sensitive design is subject to substantial legal backing in the GDPR, by means of the 'data protection by design and default (DPBD)' requirements (Article 25 GDPR). It is considered to be of the utmost importance to integrate data protection measures in the design of Big Data systems and approaches, to safeguard relevant rights and interests and to build trust in such initiatives.[15] To be able to decide on how data protection principles should be integrated in a data-intensive health research initiative, the nature, scope, context, purposes and risks related to the processing need to be taken into account, by means of a 'data protection impact assessment' (DPIA). Such a DPIA process will usually be mandatory in data-intensive health research initiatives, according to standard as set out in Article 35(1) GDPR.[16] Strategies to incorporate DPBD in Big Data initiatives include decentralised analytics models that only securely access data that is needed, mechanisms for automated policy definition and enforcement, and the further implementation and development of privacy enhancing tools.[15] Achieving DPBD in Big Data initiatives is, however, not considered to be an easy task, and a great deal of research needs to be done.[15] Indeed, some argue that the idea of integrating data protection norms in the design of information processing systems is at odds with the dynamic nature of many of these norms.[17] Nevertheless, novel solutions like DataSHIELD show that value-sensitive design can mitigate many of the normative challenges in facilitating data access and reuse.[18,19] In DataSHIELD, this is done by facilitating the co-analysis of individual-level data from multiple studies without physically sharing the data. Both the potential and the limits of DPBD or other value-sensitive design approaches in fostering legal compliance therefore need to be recognised.

Even when sufficient steps have been taken to ensure legal compliance, there still is no guarantee that all relevant interests are taken into account.[10,20] After all, legal compliance is only one of the first steps towards a responsible and proportionate approach to the governance of data-intensive health research. Mere adherence to the law does not promote or safeguard all morally relevant interests, nor does it secure public trust or a 'social license' for health research.[21] Therefore, measures beyond those that foster legal compliance are required.

In favour of an ongoing and proportionate duty to notify

The special status of scientific research in the GDPR, as we discussed above, comes with a responsibility. The many research exemptions and derogations can only be invoked, when sufficient technical and organisational measures are in place to protect data subjects, in accordance with Article 89(1) GDPR. These measures are of great importance, to compensate for the weakened protection as a consequence of the many research exemptions. This is particularly true in health research where special categories of personal data are processed, such as health-related or genetic data. Especially in this context, it should be prevented that the research exemptions from (specific) informed consent and individual rights endanger respect for the right to privacy and result in a lack of control by data subjects. Although individuals are willing to accept that they are offered a lower level of control than they would prefer, this acceptance seems to be dependent of what alternative safeguards exist to protect their interests.[22] Alternative safeguards valued by respondents in a qualitative pilot study include independent oversight and transparency.[22]

The GDPR, however, only provides limited points of departure for determining which specific measures should be in place in research.[3,7] The GDPR indicates that those measures may include pseudonymisation. What is more, a common recommendation in guidelines and the literature is that data access committees or ethics committees should be installed to review (some) of the requests for data access. Although such measures are highly advisable, they do not secure any form of transparency towards data subjects, let alone that they are enabled to control the use of their data in any way. To some extent, Patient and Public Involvement (PPI) approaches could be deployed to ensure the engagement of (representatives of) participants or data subjects,[23] but most of its modus operandi do not secure transparency towards large groups of individuals. I argue that more attention should be devoted to measures that aim to inform and involve individuals, especially when consent is not obtained or the use of data is based on a one-off broad consent procedure.

When consent is not obtained or broad consent has already been obtained in a one-off event, transparency towards data subjects is in particular promoted in the GDPR by means of the right to information in Articles 13 and 14 GDPR.[7] In these cases, the right to information also functions as a threshold for enabling data subjects to control or be involved with the specific use of their data. After all, without a notification or other knowledge about the (re)use of their data in research, individuals will not be able to exercise any of their rights, simply because they are not aware. Notifying and informing individuals, about both the reasons for

not obtaining (re)consent, and the safeguards that exist to protect their interests, may be essential to adhere to the principle of transparency, respect autonomy and maintain public trust. To advance the (reasonable) acceptance of data use without (re)consent, we propose that the information should include: 1) the reasons for not obtaining (re)consent; 2) information about the (governance) measures which could advance the acceptability for the individual of not obtaining (re)consent, and; 3) minimal but adequate information about the initiative or group of initiatives, such as the risks and benefits, and the rights of the individual. This information can be provided alongside with or in addition to any legally required information. The exact information provided to individuals should be tailored to the nature, scope, context, purposes and risks related to the use of their data. More comprehensive information should be made available on demand to prevent an information overload.

Imposing such a duty on researchers to notify individuals may, however, involve negative effects on research, such as higher costs, lower participation rates or selection bias. To avoid negative impacts on health research that cannot be justified, the duty to notify should be proportionate in relation to the possibility and practicability of notifying individuals, and the potential impact on the privacy of the individual. When it is impossible or requires a disproportionate effort to inform individuals, I suggest that the information mentioned above should still be made publically available in an easily accessible form. This, again, in addition to any legally required public information, which needs to be made available on the basis of Article 14 (5) (b) GDPR. In this proportionality test, preferably executed by an ethics committee, it should be taken into account that a duty to notify may not require much effort as long as (safe) digital solutions are taken into consideration. In addition, enhancing transparency might (on the long term) advance research data access, as a result of an improvement of trust and involvement by individuals and the public.

Conclusion

A thorough understanding of both the law and the practice of data-intensive health research are essential to allow for a flexible and proportionate approach to data sharing and access. Especially in international collaborations, even within the EU, a lack of investments in such understanding could result in overly defensive approaches to data access and sharing. The goal of the GDPR is not to prohibit but to protect personal data processing to make health research possible in the public interest. Consequently, there is no need for a prohibitive approach to data sharing or access

in health research according to the GDPR. Such a non-prohibitive approach does, however, require the implementation of a robust set of organisational and technical measures, to safeguard individual rights and interests and secure public trust. These safeguards should not merely compensate for the loss of control by data subjects by introducing alternative measures related to the principles of accountability, security and data minimisation. In addition, data holders should be bound to an ongoing and proportionate duty to inform individuals, also when there is no obligation to obtain (specific) informed consent.

References

1. Mittelstadt BD, Floridi L. The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. Sci Eng Ethics 2016; 22: 303–41.
2. Ploem MC, Essink-Bot ML, Stronks K. Proposed EU data protection regulation is a threat to medical research. BMJ 2013; 346: f3534.
3. Mostert M, Bredenoord AL, Biesaart MCIH, van Delden JJM. Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach. Eur J Hum Genet 2016; 24: 956–60.
4. Sethi N, Laurie GT. Delivering proportionate governance in the era of eHealth. Med Law Int 2013; 13: 168–204.
5. Kaye J, Whitley EA, Lund D, Morrison M, Teare H, Melham K. Dynamic consent: a patient interface for twenty-first century research networks. Eur J Hum Genet 2015; 23: 141–6.
6. Budin-Ljøsne I, Teare HJA, Kaye J et al. Dynamic Consent: a potential solution to some of the challenges of modern biomedical research. BMC Med Ethics 2017; 18: 4.
7. Mostert M. A new regulatory landscape for Big Data health research: Safeguards and research exemptions in the GDPR. (see chapter 4 of this thesis)
8. BBMRI-ERIC. A Code of Conduct for Health Research. Available at http://code-of-conduct-for-health-research.eu/.
9. Bredenoord AL, Mostert M, Isasi R, Knoppers BM. Data sharing in stem cell translational science: policy statement by the International Stem Cell Forum Ethics Working Party. Regen Med 2015; 10: 857–61.
10. Mostert M, Koomen BM, van Delden JJM, Bredenoord AL. Privacy in Big Data Psychiatric and Behavioural Research: a Multiple-Case Study. International Journal of Law and Psychiatry 2018; 60: 40-4.
11. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016. Available at https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:32016R0679.
12. Council for International Organizations of Medical Sciences (CIOMS). International Ethical Guidelines for Health-related Research Involving Humans. 2016. Available at https://cioms.ch/shop/product/international-ethical-guidelines-for-health-related-research-involving-humans/
13. World Medical Association (WMA). Declaration of Taipei on Ethical Considerations Regarding Health Databases and Biobanks. 2016. Available at https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks/
14. Voigt P, von dem Bussche A. Enforcement and Fines Under the GDPR. In: The EU General Data Protection Regulation (GDPR). Springer International Publishing: Cham, 2017, pp 201–217.

15. European Union Agency for Network and Information Security (ENISA). Privacy by design in big data. 2015. Available at https://www.enisa.europa.eu/publications/big-data-protection.
16. Article 29 Working Party. Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is 'likely to result in a high risk' for the purposes of Regulation 2016/679. 2017.
17. Koops B-J, Leenes R. Privacy regulation cannot be hardcoded. A critical comment on the 'privacy by design' provision in data-protection law. Int Rev Law, Comput Technol 2014; 28: 159–71.
18. Budin-Ljøsne I, Burton P, Isaeva J et al. DataSHIELD: An Ethically Robust Solution to Multiple-Site Individual-Level Data Analysis. Public Health Genomics 2015; 18: 87–96.
19. Wallace SE, Gaye A, Shoush O, Burton PR. Protecting Personal Data in Epidemiological Research: DataSHIELD and UK Law. Public Health Genomics 2014; 17: 149–57.
20. Kalkman S, Mostert M, van Thiel GJ, van Delden JJM. Responsible data sharing in international health research: a review of principles and norms. (see chapter 6 of this thesis)
21. Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care.data ran into trouble. J Med Ethics 2015; 41: 404–9.
22. Taylor MJ, Taylor N. Health research access to personal confidential data in England and Wales: assessing any gap in public attitude between preferable and acceptable models of consent. Life Sci Soc policy 2014; 10: 15.
23. Rose D. Patient and public involvement in health research: Ethical imperative and/or radical challenge? J Health Psychol 2014; 19: 149–158.

Appendices

Summary

In Big Data health research, researchers more comprehensively capture data and built new data analysis methods to extract valuable information from these data. The purpose of Big Data health research is not only to evaluate human generated hypotheses or answer existing questions. Big Data analytics is also deployed to recognise patterns in data that aid in hypothesis generation and raising relevant questions. What is needed for Big Data in health research to become transformative, is the wide scale collection, reuse and linkage of data, usually at the individual person level. Health research is increasingly becoming a *data-intensive* activity, in which health-related, genomic and other data about individuals are captured, reused and linked on a massive scale.

Key normative concerns related to Big Data health research relate to privacy, data protection, informed consent, anonymisation, ownership, (the myth of) objectivity, epistemology and Big Data divides or inequalities. The legislative activity in the European Union (EU) and the main focus in this thesis are on privacy and data protection. The need for a more consistent and comprehensive protection of personal data was recognised in the EU, and the Data Protection Directive 95/46/EC was replaced by the General Data Protection Regulation (GDPR). During the adoption of the GDPR many stakeholders feared that the regulation would severely restrict health research. The final version of the GDPR is not as strict as initially feared. A high level of harmonisation of the rules related to scientific research was, however, not achieved in the GDPR. It remains a challenge to interpret, balance, implement and harmonise the principles and rules in the GDPR related to health research. Moreover, the GDPR is still largely based on the same principles as Directive 95/46/EC, many of which are at odds with Big Data approaches. The large-scale reuse and linkage of personal data seem difficult to reconcile with data protection principles like purpose limitation, storage limitation and data minimisation. What is more, the discussion about the limits of informed consent and anonymity in safeguarding and balancing relevant interests remains equally important under the final GDPR. Another topic of discussion is related to a change in the key fundamental rights on which the GDPR is based, which could aid in its sound implementation and interpretation.

Against this background of normative complexity and change, stakeholders engage with the challenge of utilising the potential of Big Data in health research. The main aim of this thesis is to inform the debate about what form laws, regulations and information governance should take in the EU, to allow for progress in data-

intensive health research while safeguarding (fundamental) rights and morally relevant interests. To achieve this aim, this thesis addresses the central question of how relevant rights and interests can be safeguarded and balanced in the EU, without disproportionately hampering data-intensive health research.

This thesis starts out with a reflection on two of the key fundamental rights that underpin data protection law; the right to privacy and the right to data protection. The right to privacy has usually been considered as the most prominent fundamental right to protect in data-intensive health research. Within the EU, however, the right to data protection is gaining relevance as a separate fundamental right that should in particular be protected by data protection law. **Chapter 2** shows that there are multiple differences between these two rights, which are relevant to data-intensive health research. To begin with, the rights based on the right to data protection are of a less context-sensitive nature and easier to enforce. Then, the positive obligation to protect personal data requires a more proactive approach by the EU and its Member States. Finally, it guarantees a more comprehensive system of personal data protection. Related to health research, such a comprehensive system of data protection should be considered to serve two functions in particular. Firstly, the aim is to provide effective overarching safeguards that secure the rights and interests of individuals, irrespective of whether the personal data processing is grounded on consent or any other legal basis. The overarching safeguards should include requirements of accountability subject to independent oversight, transparency towards data subjects and the public, ensuring that data subjects can invoke their rights, and data security. Secondly, the system of data protection arranges for specific exemptions and safeguards related to data processing in scientific research. These specific safeguards should compensate for the loss of individual control as a result of the research exemptions from consent requirements, certain general principles, and individual rights.

During the reform of EU data protection law, there was a lively debate about informed consent requirements related to scientific research. To contribute to this debate, **chapter 3** reviews how the consent or anonymise approach is challenged in a data-intensive health research context, and what possible ways forward are within the EU legal framework on data protection. The 'consent or anonymise approach' has been dominant in many health research initiatives. According to this approach, only two options are available when personal health data are to be shared or accessed for research purposes. The first option is that the data are processed on the basis of informed consent. The second option is that extensive de-identification

measures are taken so that the shared data can no longer be considered personal data. Potential solutions to data sharing or access problems can therefore, according to the consent or anonymise approach, only be found by modifying or tweaking concepts of informed consent or anonymity. Although there is value in thinking about how the concept or process of informed consent or anonymisation could be improved, it should be prevented that other legal bases than informed consent are disregarded beforehand. Moreover, an overly strong focus on requirements of consent or anonymity could distract the attention from the debate on and implementation of other essential measures to protect, promote and balance relevant rights and interests. A data-intensive health research practice based on consent exemptions, accompanied with appropriate safeguards, should be preferred above continuing a practice of stretching concepts of consent or anonymisation beyond their limits.

After years of uncertainty about how the GDPR would impact data-intensive health research, the EU legislative bodies adopted the final version of the GDPR in May 2018. From this date, the possibilities of linking and analysing personal data in research are largely determined by the General Data Protection Regulation (GDPR), and by its interpretation and implementation in national law. **Chapter 4** examines whether the GDPR has achieved its dual objective of both facilitating health research and subjecting it to appropriate safeguards. After a thorough evaluation of the final version of the GDPR, I conclude that overall the GDPR does contribute to a responsible and effective use of personal data in data-intensive health research. The many open norms in the GDPR sufficiently allow for a proportionate data protection regime that is specifically tailored to the context of health research. Moreover, the large number of research exemptions and derogations combined with a specific regime of conditions and safeguards could prevent that research is disproportionately hampered. Nevertheless, it should be noted that the many open norms, and the research exemptions that need to be implemented in national law, do not contribute to a high level of harmonisation. This could negatively impact a coherent implementation and interpretation of the GDPR on the national level, and therefore potentially hinder international data sharing and access. Such a negative impact could partially be mitigated by formally recognised international codes of conduct conform Article 40 GDPR.

Chapter 5 presents the results of the multiple-case study. By means of this qualitative study, insight is gained in how privacy and data protection concerns are currently dealt with in two real world examples of Big Data health research. For this multiple-case study, the YOUth cohort, a longitudinal cohort focusing on

psychosocial development, and Big Data Psychiatry, a pilot study in Big Data analytics on psychiatric health data, were selected. Three themes emerged from the analysis of the interviews and other data. The first theme concerns the move away from anonymisation as a strategy to prevent the applicability of data protection law. Alternative measures have been implemented with the aim of complying with the law and mitigating risks. The second theme relates to the search for meaningful and proportionate ways to allow individuals to control and be aware of the use of their data. This search explores the possible use of digital tools to ensure a more ongoing engagement with participants. Thirdly, uncertainty about how to comply with the law is perceived as having negative impact on the initiatives. This uncertainty is primarily attributed to a lack of easily accessible guidance and to the recent changes in data protection law. Overall, the findings show that it takes considerable effort to take privacy and data protection norms into account in a Big Data health research initiative, especially when individual participant level data need to be linked or enriched. By embracing the complexity of the law in an early phase, setbacks could be prevented, the existing flexibility within the law could be utilised, and systems or organisations could be designed and constructed to take relevant rules into account. This chapter illustrates that a close collaboration of experts with different backgrounds within the initiative may be necessary to be able to successfully navigate this process.

In **chapter 6**, the results of a review of ethical guidelines, policy documents and literature sources for ethical principles and norms pertaining to data sharing for international health research are presented. The aim is to identify a set of ethical principles and norms to govern responsible data sharing for international health research. We observed an abundance of principles and norms with considerable convergence at the aggregate level of four overarching themes: societal benefits and value; distribution of risks, benefits and burdens; respect for individuals and groups; and public trust and engagement. However, at the level of principles and norms we identified substantial variation in the phrasing and level of detail, the number and content of norms considered necessary to protect a principle, and sometimes even contradiction between norms. Though providing some helpful leads for further work on a coherent governance framework for data sharing, the current collection of norms and principles is still too haphazard, non-uniform and sometimes even contradictory to serve as sufficient guidance in itself. Our work highlights the need for considerable investments and expertise to further develop and implement a governance framework for international data sharing projects.

More specifically, there is scarce guidance on how to deal with the limitations to preserve anonymity and confidentiality of personal data. A common recommendation is to inform participants and the public about these limitations. Future work should concentrate more on measures to safeguards rights and interests and secure public trust, at all levels of (de-)identification.

Chapter 7 focuses on data sharing in the context of stem cell science and discusses specific policy principles for responsible data sharing in stem cell translational science. In the field of today's stem cell science, the sharing of genomic- and health-related data for biomedical research is of the utmost importance. Data and sample sharing constitute a scientific and ethical imperative, but need to be conducted in a responsible manner. In 2014, the Global Alliance for Genomics and Health (GA4GH) adopted a common Framework for Responsible Sharing of Genomic and Health-Related Data. In this chapter consisting of a policy statement, this Framework is further interpreted and specified to provide guidance for the specific context of translational stem cell science. Some of the key principles need a tailored interpretation, including engagement, data quality and safety, privacy, security and confidentiality, risk-benefit analysis and sustainability. Researchers are encouraged to integrate this policy in their guidelines and protocols, in order to stimulate responsible innovation.

Finally, **chapter 8** reflects on the main findings in this thesis and discusses ways forward. I conclude that a thorough understanding of both the law and the practice of data-intensive health research are essential to allow for a flexible and proportionate approach to data sharing and access. Especially in international collaborations, even within the EU, a lack of investments in such understanding could result in overly defensive approaches to data access and sharing. The goal of the GDPR is not to prohibit but to protect personal data processing to make health research possible in the public interest. Consequently, there is no need for a prohibitive approach to data sharing or access in health research according to the GDPR. Such a non-prohibitive approach does, however, require the implementation of a robust set of organisational and technical measures, to safeguard individual rights and interests and secure public trust. These safeguards should not merely compensate for the loss of control by data subjects by introducing alternative measures related to the principles of accountability, security and data minimisation. In addition, data holders should be bound to an ongoing and proportionate duty to inform individuals, also when there is no obligation to obtain (specific) informed consent.

Samenvatting

Door de opkomst van Big Data in de medische wetenschap, worden steeds meeromvattende data verzameld en ontwikkelen onderzoekers nieuwe methoden om waardevolle informatie af te leiden uit deze data. Het doel van Big Data onderzoek is niet alleen om door mensen gegenereerde hypotheses te evalueren of bestaande vragen te beantwoorden. Big Data analyse wordt ook ingezet om patronen te herkennen in data die bij kunnen dragen aan het genereren van hypotheses en het stellen van relevante vragen. Een noodzakelijke voorwaarde voor Big Data analyse in de medische wetenschap is een grootschalige verzameling, hergebruik en koppeling van data, doorgaans op het niveau van individuele personen. De medische wetenschap wordt daarom in toenemende mate een data-intensieve activiteit, waarin gegevens over gezondheid, genetische gegevens en andere data over personen worden opgeslagen, hergebruikt en gekoppeld op een grote schaal.

De belangrijkste normatieve zorgen over de inzet van Big Data in de medische wetenschap relateren aan privacy, gegevensbescherming, geïnformeerde toestemming, anonimisatie, eigendom, (de mythe van) objectiviteit, epistemologie en verdeeldheid of ongelijkheden. De ontwikkelingen op het vlak van wetgeving in de Europese Unie (EU) en de belangrijkste focus in dit proefschrift richten zich op privacy en gegevensbescherming. De noodzaak voor een meer consistente en omvattende bescherming van persoonsgegevens werd erkend in de EU, waarna Richtlijn 95/46/EG werd vervangen door de Algemene Verordening Gegevensbescherming (AVG). Tijdens het wetgevingsproces rond de AVG vreesden veel belanghebbenden dat de verordening de medische wetenschap ernstig zou belemmeren. Uiteindelijk bleek de definitieve versie van de AVG niet zo strikt als aanvankelijk werd gevreesd. Een hoog niveau van harmonisatie van de normen, die relevant zijn voor de medische wetenschap, werd echter niet bereikt in de AVG. Het blijft mede daardoor een uitdaging om deze normen te interpreteren, balanceren, implementeren en harmoniseren. Daar komt bij dat de AVG grotendeels is gebaseerd op dezelfde uitgangspunten en principes als Richtlijn 95/46/EG, waarvan vele op gespannen voet staan met de inzet van Big Data. Het grootschalige hergebruik en koppelen van persoonsgegevens lijkt moeilijk in overeenstemming te brengen met principes als doelbinding, opslagbeperking en minimale gegevensverwerking. Ook blijft onder het regime van de AVG de discussie rond de beperkingen van het toestemmingsvereiste en anonimiteit in het beschermen en balanceren van relevante rechten en belangen onverminderd van belang. Een ander onderwerp van discussie is gerelateerd aan een verandering in de fundamentele

rechten waarop de AVG in het bijzonder op is gebaseerd, die van belang is voor een juiste interpretatie en implementatie van de verordening.

Tegen deze achtergrond van normatieve complexiteit en verandering, gaan onderzoekers en andere belanghebbenden de uitdaging aan om het potentieel van Big Data in de medische wetenschap te benutten. Dit proefschrift heeft als doel om het normatieve debat te verrijken over hoe vooruitgang in de data-intensieve medische wetenschap samen kan gaan met de bescherming van (fundamentele) rechten en moreel relevante belangen. Om dit doel te bereiken, adresseert dit proefschrift de hoofdvraag hoe relevante rechten en belangen beschermd en gebalanceerd kunnen worden in de EU, zonder dat de data-intensieve medische wetenschap disproportioneel belemmerd wordt.

Dit proefschrift begint met een reflectie op twee fundamentele rechten die ten grondslag liggen aan gegevensbescherming in de EU; het recht op privacy en het recht op gegevensbescherming. Het recht op privacy wordt van oudsher gezien als het belangrijkste recht dat beschermd moet worden in de data-intensieve medische wetenschap. In de EU neemt het recht op gegevensbescherming echter een steeds belangrijkere plek in als apart fundamenteel recht, dat in het bijzonder gewaarborgd wordt in de AVG. **Hoofdstuk 2** laat zien dat er meerdere verschillen zijn tussen beide fundamentele rechten die relevant zijn voor de data-intensieve medische wetenschap. Rechten gebaseerd op het fundamentele recht op gegevensbescherming zijn van een minder context-afhankelijke aard en beter te handhaven. Voorts vereist de positieve verplichting om persoonsgegevens te beschermen een meer proactieve aanpak door de EU en haar lidstaten. Ten slotte garandeert het een meeromvattend systeem van normen die persoonsgegevens beschermen. In verhouding tot de medische wetenschap, vervult een dergelijk systeem van gegevensbescherming twee functies in het bijzonder. Ten eerste heeft het tot doel om effectieve overkoepelende waarborgen te bieden die de rechten en belangen van individuen beschermen, ongeacht of de verwerking van persoonsgegevens is gebaseerd op een geïnformeerde toestemming of een andere legitieme grondslag. Deze overkoepelende bescherming moet onder meer bestaan uit vereisten rond het afleggen van verantwoording, het bieden van transparantie richting betrokkenen en het publiek, ervoor zorg dragen dat betrokkenen hun rechten kunnen uitoefenen en het beveiligen van persoonsgegevens. Ten tweede bevat het systeem van gegevensbescherming specifieke uitzonderingen en waarborgen voor gegevensverwerking in wetenschappelijk onderzoek. Deze specifieke waarborgen moeten compensatie bieden voor het verlies aan controle

door de betrokkene als gevolg van de uitzonderingen op toestemmingsvereisten, sommige kernprincipes en individuele rechten.

Gedurende de hervorming van de EU-regels rond gegevensbescherming ontstond een levendig debat rond het toestemmingsvereiste. Om bij te dragen aan dit debat, biedt **hoofdstuk 3** inzicht in hoe een benadering waarin geïnformeerde toestemming wordt gevraagd of data worden geanonimiseerd onder druk staat in de context van data-intensieve medische wetenschap. Daarnaast verkent dit hoofdstuk mogelijke oplossingen en alternatieven binnen het juridische kader van gegevensbescherming in de EU. De 'toestemming of anonimiseren' benadering is dominant geweest bij veel initiatieven in de medische wetenschap. Volgens deze benadering zijn er slechts twee mogelijkheden om gegevens over personen te gebruiken. De eerste mogelijkheid is dat persoonsgegevens verwerkt worden op basis van een geïnformeerde toestemming. De tweede optie is dat de-identificatie maatregelen worden genomen, zodat de gegevens niet langer als persoonsgegevens worden beschouwd. Gevangen in een dergelijke toestemming of anonimiseren dichotomie, kunnen potentiele oplossingen voor problemen rond het delen of hergebruiken van data alleen gevonden worden in het sleutelen aan concepten van geïnformeerde toestemming en anonimiteit. Hoewel het waardevol is om na te denken over hoe het vragen van toestemming of anonimiseren van data verbeterd kan worden, moet voorkomen worden dat andere juridische grondslagen dan toestemming bij voorbaat buiten beschouwing worden gelaten. Bovendien schuilt in een overmatig sterke focus op vereisten rond toestemming of anonimiteit het gevaar dat het debat rond en invoering van andere essentiële maatregelen om rechten en belangen te beschermen en balanceren in het gedrang komen. Een praktijk van data-intensieve medische wetenschap gebaseerd op een uitzondering op het toestemmingsvereiste, onderworpen aan passende voorwaarden en waarborgen, moet verkozen worden boven het oneigenlijk ver oprekken van concepten van geïnformeerde toestemming of anonimiteit.

Na jaren van onzekerheid over wat de impact van de AVG op de medische wetenschap zou zijn, werd de AVG in mei 2018 aangenomen. Vanaf dat moment werd het (internationale) juridische speelveld voor het hergebruiken en koppelen van persoonsgegevens in de data-intensieve medische wetenschap grotendeels bepaald door de AVG. Ook is de interpretatie en implementatie van de AVG op nationaal niveau van belang. In **hoofdstuk 4** wordt onderzocht of de AVG de dubbele doelstelling heeft behaald om medisch-wetenschappelijk onderzoek met persoonsgegevens te faciliteren én dit onderzoek te onderwerpen aan passende waarborgen. Na een grondige analyse van de AVG, kom ik tot de conclusie dat de AVG bijdraagt aan

een verantwoord en effectief gebruik van persoonsgegevens in de data-intensieve medische wetenschap. De veelal open normen in de AVG bieden voldoende ruimte voor een proportioneel regime van gegevensbescherming dat toegesneden is op de medische wetenschap. Hierbij zijn de combinatie van uitzonderingen en afwijkingen ten behoeve van wetenschappelijk onderzoek met het vereiste dat in passende en specifieke waarborgen wordt voorzien van essentieel belang. Hierbij past wel de kanttekening dat vele van de uitzonderingen voor wetenschappelijk onderzoek geïmplementeerd moeten worden in nationale wetgeving van lidstaten, waardoor de centrale doelstelling van harmonisatie in de EU beperkt is bereikt. Het risico is dan ook groot dat het negatieve effect op de data-intensieve medische wetenschap als gevolg van het gebrek aan een coherente bescherming in de EU voorlopig blijft bestaan. Deze negatieve impact kan wel verminderd worden door de erkenning van internationale gedragscodes op grond van artikel 40 AVG, specifiek toegesneden op de medische wetenschap.

In **hoofdstuk 5** worden de resultaten van de meervoudige casestudy besproken. Door middel van dit kwalitatieve onderzoek is inzicht verkregen in hoe omgegaan wordt met zorgen rond privacy en gegevensbescherming in twee voorbeelden van Big Data onderzoek. De cases in deze studie betreffen het YOUth cohort, een longitudinaal cohort dat zich richt op psychosociale ontwikkeling, en Big Data Psychiatrie, een pilot studie in Big Data analyse met gegevens over psychiatrische patiënten. Uit de analyse van de afgenomen interviews en andere verzamelde data kwamen drie thema's naar voren. Het eerste thema betreft het afstand nemen van een strategie waarbij gegevens worden geanonimiseerd om te voorkomen dat de wettelijke bescherming van persoonsgegevens van toepassing is. Alternatieve maatregelen worden genomen met het doel om aan de wet te voldoen en risico's te verminderen. Het tweede thema is gerelateerd aan de zoektocht naar betekenisvolle en proportionele manieren om individuen in staat te stellen om bewust te zijn van het gebruik van hun data en hier controle over uit te oefenen. In deze zoektocht wordt de potentie van digitale tools verkend om een meer voortdurende betrokkenheid van deelnemers te bewerkstelligen. Ten derde bestaat veel onzekerheid over juridische normen waardoor de voortgang in de initiatieven negatief beïnvloed wordt. Deze onzekerheid wordt vooral toegeschreven aan een gebrek aan toegankelijke richtlijnen en aan de recente wijzigingen in de wetgeving inzake gegevensbescherming. Thema overstijgend laten de bevindingen zien dat het aanzienlijke inspanningen vergt om normen rond privacy en gegevensbescherming in acht te nemen, in het bijzonder wanneer gegevens over individuen gekoppeld of verrijkt moeten worden. Door de

complexiteit van het recht in een vroege fase te omarmen, kunnen tegenslagen voorkomen worden, kan de bestaande flexibiliteit in het recht worden benut en kunnen organisaties en systemen ontwikkeld worden met inachtneming van relevante normen. Dit hoofdstuk illustreert dat een nauwe samenwerking tussen experts met verschillende achtergronden essentieel lijkt om een medisch-wetenschappelijk Big Data initiatief op te zetten.

In **hoofdstuk 6**, komen de resultaten van een review van richtlijnen, beleidsdocumenten en bronnen uit de literatuur over ethische principes en normen betreffende het delen van data voor internationaal gezondheidsonderzoek aan de orde. Het doel van dit hoofdstuk is om een set van ethische principes en normen te identificeren gericht op het verantwoord delen van data voor internationaal medisch-wetenschappelijk onderzoek. We vonden een overvloed aan principes en normen met een aanzienlijke convergentie op het niveau van vier overkoepelende thema's: maatschappelijke voordelen en waarde; verdeling van risico's, voordelen en lasten; respect voor individuen en groepen; en publiek vertrouwen en engagement. Op het niveau van principes en normen, vonden we echter een aanzienlijke variatie in de formulering en het detailniveau, het aantal en de inhoud van normen en principes, en soms zelfs tegenstellingen tussen normen. Hoewel ze nuttige aanknopingspunten bieden voor de ontwikkeling van een samenhangend governancekader voor gegevensuitwisseling, is de huidige verzameling normen en beginselen nog steeds te gevarieerd, niet-uniform en soms zelfs tegenstrijdig om op zichzelf als richtlijn te dienen. Ons werk benadrukt de noodzaak van aanzienlijke investeringen en expertise om een governancekader voor internationale gegevensuitwisselingsprojecten verder te ontwikkelen en te implementeren. Meer specifiek zijn er nauwelijks richtlijnen over hoe om te gaan met de beperkingen in het bewaken van anonimiteit of vertrouwelijkheid van persoonsgegevens. Een veelvoorkomende aanbeveling is om deelnemers en het publiek te informeren over deze beperkingen. Toekomstig werk moet zich meer richten op maatregelen die rechten en belangen beschermen en het vertrouwen van het publiek veilig te stellen, op alle niveaus van (de-)identificatie.

Hoofdstuk 7 richt zich op specifieke normen en beleidsprincipes voor verantwoord delen van gegevens in de translationele stamcelwetenschap. Op het gebied van de hedendaagse stamcelwetenschap is het delen van genomische en gezondheidsgerelateerde data voor biomedisch onderzoek van het grootste belang. Het delen van gegevens vormt een wetenschappelijk en ethisch uitgangspunt, maar moet op een verantwoorde manier plaatsvinden. In 2014 heeft de Global Alliance for Genomics and Health (GA4GH) een gemeenschappelijk raamwerk voor

het verantwoord delen van genomische en gezondheidsgerelateerde gegevens aangenomen. In dit hoofdstuk, bestaande uit een policy statement, wordt dit raamwerk verder geïnterpreteerd en gespecificeerd als leidraad voor de specifieke context van translationele stamcelwetenschap. Enkele van de belangrijkste principes vereisen een interpretatie op maat, inclusief de principes van engagement, gegevenskwaliteit en veiligheid, privacy, beveiliging en vertrouwelijkheid, risico-batenanalyse en duurzaamheid. Om verantwoorde innovatie te stimuleren worden onderzoekers aangemoedigd om dit beleid te integreren in hun richtlijnen en protocollen.

Tot slot, reflecteert **hoofdstuk 8** op de belangrijkste bevindingen in dit proefschrift en bespreekt het mogelijke oplossingen of benaderingen. Ik concludeer dat een grondig begrip van zowel het normatief kader rond als de praktijk van de data-intensieve medische wetenschap essentieel zijn om een flexibele en proportionele benadering van gegevensuitwisseling mogelijk te maken. Vooral in internationale samenwerkingsverbanden, zelfs binnen de EU, kan een gebrek aan investeringen in een dergelijk begrip resulteren in defensieve benaderingen van toegang tot gegevens. Het doel van de AVG is niet om de verwerking van persoonsgegevens te verbieden, maar om deze verwerking te beschermen en wetenschappelijk onderzoek met persoonsgegevens in het algemeen belang mogelijk te maken. Op grond van het gedachtengoed achter de AVG is er geen noodzaak voor een prohibitieve benadering van gegevensuitwisseling in de medische wetenschap. Een dergelijke niet-prohibitieve aanpak vereist echter wel robuuste organisatorische en technische maatregelen om individuele rechten en belangen te beschermen en het vertrouwen van het publiek te waarborgen. Deze waarborgen moeten niet alleen het verlies van controle door betrokkenen compenseren door alternatieve maatregelen in te voeren die verband houden met de beginselen van verantwoording, beveiliging en gegevensminimalisatie. Ook moeten verantwoordelijken gebonden zijn aan een voortdurende maar proportionele plicht om personen te informeren, zelfs wanneer er geen plicht bestaat tot het verkrijgen van een (specifieke) geïnformeerde toestemming.

Curriculum Vitae

Menno Mostert was born on 2 Februari 1984 in Utrecht, the Netherlands. In 2008 he obtained his master's degree in Dutch Law in Groningen, specialised in private- and criminal law. This same year, he started his training as a lawyer specialised in health law at the law firm Van Benthem & Keulen in Utrecht. After completion of his training as a lawyer, he practised law for another year at the law firm SmeetsGijbels in Amsterdam. In 2012 he started teaching health law at the Department of Medical Humanities at the Julius Center for Health Sciences and Primary Care of the University Medical Center (UMC) Utrecht. In 2014 he started his work on this thesis as a PhD student at the Julius Center, where he currently works as an assistant professor. He teaches medical students and medical doctors in the field of health law and medical humanities. Besides teaching, he contributes to the legal and ethical research of his department, currently in the BigData@Heart project of the Innovative Medicines Initiative. Furthermore, he is a deputy lawyer at the Research Ethics Committee of the UMC Utrecht and a deputy secretary at the Regional Medical Disciplinary Tribunal in Zwolle.

Dankwoord

Graag bedank ik iedereen die heeft bijgedragen aan dit proefschrift, waarvan ik de volgende personen in het bijzonder noem.

Prof. dr. J.J.M van Delden, beste Hans, veel dank voor het mogelijk maken van dit waardevolle traject en de uitstekende begeleiding. Jouw visie en de manier waarop je mij hielp om focus aan te brengen waren onmisbaar.

Prof. dr. A.L. Bredenoord, beste Annelien, het is bewonderingswaardig hoe jij ambitie en inhoudelijke scherpte weet te combineren met positiviteit en menselijkheid. Na iedere bespreking ging ik met hernieuwde energie aan de slag. Het was een voorrecht om door jou begeleidt te worden.

Mr. M.C.I.H. Biesaart, beste Monique, dank voor de begeleiding bij het eerste gedeelte van mijn promotie en de vele gezellige momenten als (oud-)kamergenoot. Als UHD gezondheidsrecht speelde je een belangrijke rol in mijn ontwikkeling en carrière als docent.

Mr. dr. Bart van der Sloot, Bregje Koomen, dr. Shona Kalkman en dr. Ghislaine van der Thiel, ieder van jullie dank voor de prettige samenwerking aan één van de manuscripten waarop dit proefschrift is gebaseerd.

Prof. mr. dr. H.D.C. Roscam Abbing, hartelijk dank voor uw belangeloze bijdrage aan het derde hoofdstuk in dit proefschrift. Het boek dat u mij toezond met de titel "Ik laat mijn hand niet sturen" staat prominent boven mijn bureau en inspireert mij om het juiste te doen.

Hooggeleerde leden van de leescommissie, prof. dr. F.E. Scheepers, prof. dr. ir. R.C.H. Vermeulen, prof. mr. dr. J.G. Sijmons, prof. dr. M.J. van den Hoven en prof. dr. D.E. Grobbee, dank voor het beoordelen van dit proefschrift.

Verder wil ik de geïnterviewden en andere betrokkenen bij Big Data Psychiatrie en het YOUth cohort bedanken voor hun tijd en inzet.

Ook veel dank aan mijn collega's en oud-collega's van de afdeling Medical Humanities. Jullie feedback en de gesprekken, onder meer tijdens Journal Clubs en OIO overleggen, waren van grote waarde.

Dan mijn paranimfen, dr. Megan Milota en Rob Schrier. Rob ken ik sinds de basisschool, Megan sinds twee jaar als zeer gewaardeerde collega en kamergenoot. Dank voor de vriendschap en jullie bereidheid om mij bij te staan als paranimf.

Een belangrijk deel van dit proefschrift is tot stand gekomen in het tuinhuis op 'Landgoed de Plaggenborgh', waar ik in de watten werd gelegd en de rust vond om

te schrijven. Aan al mijn vier ouders, dank voor jullie onvoorwaardelijke zorgen en liefde, ook als grootouders van Veere.

Saskia, ik ben dankbaar voor onze liefde en je vertrouwen in mij. Niets is mij zo dierbaar als ons gezin, met naast onze schat Veere nog een klein wonder op komst.