SODA

Scalable Oblivious Data Analytics

D3.5 Use-case specific legal aspects

Gerald Spindler (GU), Anna Zsófia Horváth (GU),

The project SODA has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 731583.

Project Information

Scalable Oblivious Data Analytics

Project number:	731583
Strategic objective:	H2020-ICT-2016-1
Starting date:	2017-01-01
Ending data:	2019-12-31
Website:	https://soda-project.eu/

Document Information

Title:	Use-case specific legal aspects				
ID:	D3.5	Type:	R	Dissemination level:	PU
		Month:	M36	Release date:	31 December 2019

Contributors, Editor & Reviewer Information

Contributors (person/partner: sections)	Gerald Spindler (GU) Anna Zsófia Horváth (GU)
Editor (person/partner)	Anna Zsófia Horváth (GU)
Reviewer (person/partner)	Paul Koster (PHI)

Release History

Release number	Date issued	SVN version	Release description / changes made
1.0	31 December 2019		First release to EU

SODA Consortium

Full Name	Abbreviated Name	Country
Philips Electronics Nederland B.V.	PHI	Netherlands
Alexandra Institute	ALX	Denmark
Aarhus University	AU	Denmark
Göttingen University	GU	Germany
Eindhoven University of Technology	TUE	Netherlands

Table 1: Consortium Members

Table of Contents

List of Figures

List of Abbreviations

AG	Advocate General
Art.	Article
Art. 29 WP	Article 29 Working Party
BayLDA	Bavarian Data Protection Supervision Agency (Bayerische Landesamt für Datenschutzaufsicht)
BGH	German Federal Court of Justice (Bundesgerichtshof)
ch	chapter
CIPL	Centre for Information Policy Leadership, Hunton Andrews Kurth LLP
CJEU	Court of Justice of the European Union
CNIL	National Commission on Informatics and Liberty (Commission Nationale Informatique & Libertés)
CR	Computer und Recht
DG	Directorate-General
DSRITB	Datenschutzrecht International-Tagungsband
DuD	Datenschutz und Datensicherheit
ECJ	European Court of Justice
EDPB	European Data Protection Board
EDPL	European Data Protection Law Review
EDPS	European Data Protection Supervisor
ENISA	European Union Agency for Cybersecurity
EU	European Union
FDPA	Federal Data Protection Act of Germany (Bundesdatenschutzgesetz, BDSG)
JIPITEC	Journal of Intellectual Property, Information Technology and E-Commerce Law
ICO	Information Commissioner’s Office
IDPL	Journal of International Data Protection Law
ISO	International Standardization Organisation
GDPR	EU General Data Protection Regulation (Regulation 2016/679)
MMR	Multimedia und Recht
MPC	Multi-Party Computation
NJW	Neue Juristische Wochenschrift
OECD	Organisation for Economic Co-operation and Development
para.	paragraph
PETs	Privacy Enhancing Technologies
R	Recital
s	sentence
TFEU	Treaty on the Functioning of the European Union
ZD	Zeitschrift für Datenschutz

Executive Summary

One objective of the SODA work package 3 is to ensure compliance with the European data protection and privacy framework. It is of special interest to provide precise legal evaluation to steer technical developments towards solutions that help in achieving this goal.

With this aim in mind, this deliverable presents an overview that includes a detailed discussion of the most relevant data protection provisions from the perspective of the project, and the analysis of some case studies useful for evaluating the techniques developed by SODA from a legal perspective.

The first part of this deliverable provides a focused overview of the legal issues concerning the analysis of health data on a large-scale using privacy-preserving, secure cryptographic protocols. It discusses carefully selected risks and challenges that are either inherent in the case studies or inevitably occur at some stage of the data processing. Anonymised data is different from anonymous data, as it implies that the data was once personal, it is not anonymous right from the onset. Therefore, even when deploying certain PETs, data protection rules apply at least at the early phase of the processing. The sensitive nature of health data, the use of innovative technologies and large-scale data processing are all important factors that need to be taken into due account when determining the particular details of the processing. The cumulative effect of more risk factors requires the careful evaluation and selection of the purposes and lawful basis of processing and technical and organisational measures. All these circumstances of data management affect the question of identifiability, how and when personal data can be considered not identifiable. One of the key contributions of this work is the tailored assessment of key provisions of the GDPR specially identified for the SODA data processing scenarios.

The second part of the document presents three case studies, one leading study and its two modified versions. In essence, these cases are very similar. Therefore, they are perfectly suitable for showing that minor changes in the data processing setting on an operational level can have serious implications from a legal standpoint. This part intends to demonstrate how cryptographic solutions such as secret sharing, multi-party computation and differential privacy have the potential to de-identify personal data in a way that does not allow the identification of the individuals. This sort of secure de-identification reduces risks and enhances privacy. On that account, it has the potential to improve the legal situation of data controllers and data processors while not compromising the rights and freedoms of the data subjects.

About this Document

Role of the deliverable

This deliverable is an application-oriented legal analysis based on in-depth legal research carried out during the second half of the SODA project. It is part of Work Package 3 of SODA and the second deliverable addressing legal and regulatory topics.

The main objective of the overall project is to enable secure MPC techniques for big data applications, and to make this vision a commonplace reality, several data protection related legal challenges need to be addressed. In line with the specific objectives of Work Package 3 to help develop GDPR compliant data analytic techniques (Objective O2.b), this deliverable presents three legal case studies and discusses use-case specific legal aspects. It is an important step towards achieving milestone 3.2 “legal analysis”.

Relationship to other SODA deliverables

This Deliverable builds upon deliverable D3.1, which examined the general legal aspects of privacy preserving Big Data analytics.

The cases discussed here are in line with the user studies presented in D3.2 ‘User studies plan’ and D3.3 ‘User studies analysis’.

Relationship to other versions of this deliverable

This deliverable only has a single version.

Structure of this document

The document consists of 4 major sections. After a brief introduction (1) the deliverable continues with an analysis of identified risks and key challenges (2) and three case studies (3). The legal assessment then ends with the conclusions and closing remarks (4).

1. Introduction

The European Commission considers data-driven innovation as a key factor for the European Union in achieving its full potential in terms of research and development.[1] The Big Data sector has been growing by 40 % per year, therefore, maximising the growth of the digital economy is rightfully one of the three pillars of the Digital Single Market Strategy.[2] Since data has become such a valuable asset, often referred to as "the oil of the 21th century", there are strong incentives for developing new and innovative solutions for the efficient utilisation of the vast amount of information and knowledge locked in the data.[3] One of the sectors that can benefit most from developments in digitalisation and datafication is the healthcare industry which is one of the largest and most rapidly growing industry sectors of the world. Enabling Big Data analytics on medical data has the potential to reduce costs and improve efficiency in public and private healthcare.[4] New insights are opening up due to the increasing linkability of data, for example through the combination of genetic data with other patient information, or even with non-sensitive personal data.

However, while allowing data exploitation for medical research purposes and innovation is extremely beneficial for the industry, one must not forget about the potential ramifications for individuals and the inherent risks of violation of fundamental rights, such as the right of informational self-determination. The adoption of the General Data Protection Regulation (thereafter GDPR) ensures that "*the processing of personal data is governed by uniform, up-to-date rules throughout the Union*".[5] These new set of rules serve a dual purpose: the protection of natural persons and the free movement of personal data within the EU.[6] Both these objectives root in the Charter of Fundamental Rights of the European Union, Art. 8 and Art. 13 respectively.[7] Consequently, it is paramount to find a fine balance between data exploitation and data protection.

The latest developments of privacy-preserving technologies might be an answer to this problem commonly referred to as the privacy-utility trade-off. Secure, privacy-preserving data analytics that are capable of large-scale processing are suitable for creating trust in digital services and improving data sharing amongst others for research purposes. For statistical and scientific research purposes that do not need personal identification the ideal way of protecting personal data would be to convert personal data to non-personal data and to carry out the actual data analysis on the anonymised dataset. Anonymisation of personal data is a complex question that the GDPR only partially answers. Anonymised data falls outside the scope of data protection, yet it remains unclear, under which circumstances data is not personally identifiable anymore. Concerns related to the question of identifiability are legitimate concerns. The wide variety of data processing settings cannot be described with a one-size-fits-all, rigid set of criteria, but it calls for individual scrutiny instead.

Generally, it can be stated that implementing multiple privacy-preserving methods that complement and strengthen one another offers high-level security. The higher security level there is, the better the protection against intruders will be. Ideally, a data processing system is robust enough to withstand

[1] European Commission, 'Towards a common European data space' (Communication) COM (2018) 232 final, 1

[2] European Commission, 'A Digital Single Market Strategy for Europe' (Communication) COM (2015) 192 final, 13-14

[3] European Commission, 'Building a European Data Economy' (Communication) COM (2017) 9 final, 1

[4] European Commission, 'Enabling the digital transformation of health and care in the Digital Single Market; empowering citizens and building a healthier society' (Communication) COM (2018) 233 final, 1-3

[5] European Commission (Communication) COM (2015) 192 final, 14

[6] R 2-7 GDPR

[7] Charter of Fundamental Rights of the European Union [2007] OJ C303/1

unlawful attempts to access and decipher the data. This is exactly the ambitious aim SODA attempts to achieve. Arguably, secret sharing and multi-party computation, combined with other strategies, such as differential privacy provide for a data processing method where once personal data is processed in an anonymised form, in technical *and* legal sense as well. In order to decide whether secret-shared or otherwise de-identified data are secure enough for the – at least partial – non-applicability of the GDPR or the prevention of unwanted identification, the legal background of identifiability as well as data processing using privacy-preserving methods must be assessed.

To that end, the primary objective of this Deliverable is to contribute to the development of a reliable and secure data processing system compliant with the provisions of the current data protection regulatory regime. After reviewing the most important data protection provisions relevant for Big Data analytics with medical data on an abstract-general level in Deliverable 3.1, it is now time to turn to certain application scenarios and follow an event-based approach. Therefore, this Deliverable focuses on the use-case specific legal aspects.

This Deliverable consists of two parts. The first part begins by examining the risks associated with the processing activities of the use cases and provides a focused overview of these. Next, it considers the most pressing data protection and privacy issues in connection with the cases. These challenges were selected based on the previously identified risks. The second part provides an accurate, in-depth legal case review of three different applications: two intra-sector and one inter-sector data processing setting. Each of these pilot cases intend to demonstrate how removing the identifiable attributes of personal data with secret sharing, MPC and differential privacy is suitable to reduce risks, enhance the level of data security and how it could even lead to anonymised data the GDPR does not apply for.

2. Identified risks and key challenges

This section discusses the risks and challenges arising from the processing of high volumes of cryptographically protected personal data related to health for scientific research and statistical purposes. Understanding these risks is of pivotal importance. It facilitates the deployment of appropriate techniques adequate to the given situation which is an important obligation of every data controller.[8] The notion of privacy by design demands that data protection functionalities shall not only be in place during the processing, but shall also be preconfigured.[9] Proactive data management incorporates both the establishment of proper contractual relations, transparency and the integration of appropriate data processing mechanisms.[10] In order to put together a legally compliant and technically feasible data processing design it is essential to map and assess the potential risk factors and to decide on the key questions and cornerstones of the processing in advance.

2.1. Identified risks

Identifying specific risks is line with the risk-based approach to data protection embraced by the GDPR.[11] This encourages stakeholders to anticipate the risk level of their data processing activities and to proactively implement protective measures corresponding to those risks.[12]
For this reason, it may at first sight appear surprising, that the GDPR does not include a definition of privacy risk. Reference can, however, be made to several other sources. ISO defines privacy risk as "*an effect of uncertainty on privacy*".[13] Uncertainty is explained as "*the state, even partial, of deficiency of information related to, understanding or knowledge of, an event, its consequence, or likelihood*".[14] Pursuant to the statement of the Art. 29 WP on the role of a risk-based approach risks are related to "*potential negative impact on the data subject's rights, freedoms and interests*".[15] The concept of impact

[8] Art. 25 GDPR, *Privacy by design*

[9] Marit Hansen, Art. 25 para 8. in Spiros Simitis, Gerrit Hornung and Indra Spiecker (eds), *Datenschutzrecht DSGVO mit BDSG* (Nomos 2019)

[10] introducing the term "contractual data anonymisation" as one of the architectural elements of Privacy by Design in the age of Big Data Ann Cavoukian, *Privacy by Design: From rhetoric to reality* (Information and Privacy Ontario, 2012) ch 3 66; Ann Cavoukian, 'Privacy by design – The 7 Foundational Principles, Implementation and Mapping of Fair Information Practices' (2012) available at https://www.iab.org/wp-content/IAB-uploads/2011/03/fred_carter.pdf accessed 26 November 2019

[11] Forum Privatheit, Datenschutz-Folgenabschätzung - Entwicklung und gegenwärtige Praxis (White Paper, Cm, 2016) 7; Winfried Veil, 'DS-GVO: Risikobasierter Ansatz statt rigides Verbotsprinzip - Eine erste Bestandsaufnahme' [2015] ZD 347, 353

[12] BayLDA, 8. Tätigkeitsbericht des Bayerischen Landesamts für Datenschutzaufsicht für die Jahre 2017 und 2018, submitted in March 2019, 130, available at https://www.lda.bayern.de/media/baylda_report_08.pdf accessed 26 November 2019

[13] ISO/IEC 29100:2011 (E) International Standard, Information technology – Security techniques – Privacy framework, 3, No. 2.19, available at: https://standards.iso.org/ittf/PubliclyAvailableStandards/c045123_ISO_IEC_29100_2011.zip accessed 26 November 2019

[14] ISO (n 13) 3, No. 2.19 Note 2

[15] Art. 29 WP, Statement on the role of a risk-based approach in data protection legal frameworks [2014] 14/EN WP 218, 4, available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp218_en.pdf accessed 26 November 2019

is generally broadly understood to refer to both tangible and objectively assessed non-tangible harm.[16] Such negative impact may result from situations like illegitimate access to personal data, unintended change in personal data or disappearance of personal data.[17]

The criteria that determine the dimensions of the impact and help quantify the risk level are called risk factors. According to R 75 GDPR these risk factors are, inter alia, potential discrimination or loss of confidentiality of personal data protected by professional secrecy, the sensitive nature of the data, evaluation of personal aspects such as health, vulnerability of the data subjects or large-scale data processing.

When determining the likelihood and severity of the risk for the overall processing pursuant to Art. 35 (1) GDPR account should be given to the nature, scope, context and purposes of the processing.[18] Although the potential harm associated with the data processing is contextual, the risk evaluation should always be conducted based on an objective assessment.[19] Pursuant to R 76 processing operations are then categorised involving either risk or high risk. The severity of risks shifts on a scale from minimal/low risk to high risk, depending on the risk factors involved. The more there are, the higher the risks of a given processing operation will be.[20] This distinction between different levels of risks based on the severity establishes different obligations and consequences for the controllers and processors.

According to the risk-based approach, processing of sensitive data, processing activities that affect vulnerable individuals as well as large-scale processing are per se processing operations associated with certain level of risk. Cumulating these factors will result in “high-risk” processing.[21]

Against this background, the following risk factors are identified in the SODA cases:[22]

[16] OECD, Supplementary Explanatory Memorandum to OECD Guidelines C(80)58/FINAL, 24, available at: http://www.oecd.org/sti/ieconomy/2013-oecd-privacy-guidelines.pdf ; CIPL, The Role of Risk Management in Data Protection (White Paper, Cm, 2014) 17, available at https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/white_paper_2-the_role_of_risk_management_in_data_protection-c.pdf accessed 26 November 2019

[17] CNIL, ‘Methodology for Privacy Risk Management’[2012] 6, available at www.cnil.fr/sites/default/files/typo/document/CNIL-ManagingPrivacyRisks-Methodology.pdf accessed 26 November 2019

[18] R 76 s 1; Datenschutzkonferenz, ‘Datenschutz-Folgenabschätzung nach Art. 35 DS-GVO‘ (2018) 5 DSK-Kurzpapier 1, 3, available at https://www.datenschutzkonferenz-online.de/media/kp/dsk_kpnr_5.pdf accessed 26 November 2019

[19] BayLDA Tätigkeitsbericht zur Sicherheit der Verarbeitung – Art. 32 DSGVO, 9 July 2016, 2, available at https://www.lda.bayern.de/media/baylda_ds-gvo_1_security.pdf accessed 26 November 2019

[20] R 76 s 2

[21] Silke Jandt, Art. 35 para. 7. in Jürgen Kühling Benedikt Buchner (eds), *Kommentar zur Datenschutz – Grundverordnung/BDSG* (2nd edn, C.H.Beck 2018); Art. 29 WP, 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, 17/EN, WP 248rev.01 7, available at https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236
accessed 26 November 2019

[22] The taxonomy is based on Art. 29 WP, WP 248 (n 21) 9 f

2.1.1. Processing of special categories of personal data

These are sensitive data of highly personal nature. Art. 9 (1) defines special categories of personal data as *“personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person,* ***data concerning health*** *or data concerning a natural person's sex life or sexual orientation, and explicitly prohibits the processing of those data”*.
According to Art. 4 No. 15 health data means *“personal data related to the physical or mental health of a natural person, including the provision of health services, which reveal information about his or her health status”*. This is a much broader definition than the term medical data; it encompasses all kind of information that may reveal details about one’s general health conditions.[23] R 35 lists “*information derived from the testing or examination of a body part or bodily substance, including from genetic data and biological samples; and any information on, for example, a disease, disability, disease risk, medical history, clinical treatment or the physiological or biomedical state of the data subject independent of its source, for example from a physician or other health professional, a hospital, a medical device or an in vitro diagnostic test*” as examples for health data. Furthermore, medical records, data on medication and diseases, a person’s intellectual and emotional capacity, smoking or drinking habits, allergies disclosed to public bodies or private entities but even the membership in a patient support group are all data concerning the health of individual data subjects belong to this group as well.[24] Besides, it also includes information on time spent in hospital, health or wellness clinic, data exchanged between patient and caretaker or doctor via telecommunication services, browser history, downloaded smartphone or wearables applications as well as other relevant data emerging from the “Internet of Things”.[25] Unlocking Big Data, new forms of analytics and systematic data mining technologies challenge the ability of differentiating between sensitive data and regular personal data. Health data can be deducted from a number of datasets, such as financial claim data, banking data, or bank account statement.[26]

2.1.2. Processing of data protected by instruments other than the GDPR

Patient records and medical data are strongly protected not only by the General Data Protection Regulation but also by additional soft law tools, recommendations and legal instruments of professional organisations, ethical guidelines, etc.
First and foremost, medical data is subject to medical confidentiality and patient-doctor privilege and is therefore protected not only by general data protection regulations but

[23] Britta Alexandra Mester, Art. 9 para. 15 in Jürgen Taeger and Detlev Gabel (eds), *Kommentar DSGVO – BDSG* (3th edn, Fachmedien Recht und Wirtschaft – dfv Mediengruppe, 2019); Christina Kühnl and others, ‘Ein europäischer Gesundheitsdatenschutz’ [2018] DuD 735, 737

[24] Art. 29 WP, Health data in apps and devices, Annex to the response of the Art. 29 WP to DG Connect (Mr. Timmers), 2, available at: https://ec.europa.eu/justice/article-29/documentation/other-document/files/2015/20150205_letter_art29wp_ec_health_data_after_plenary_annex_en.pdf

[25] Thilo Weichert, Art. 9 para. 39, in Kühling and Buchner (eds) (n 21)

[26] Tal Z. Zarsky, ‘Incompatible: The GDPR in the Age of Big Data’ (2017) 47/2 Seton Hall Law Review 995, 1013 argues, that Big Data eventually undermines the distinction between different categories of personal data. On the other hand, he accepts the need for enhanced protection of specific categories of personal data.

also by professional secrecy. The duty of confidentiality protects the right to informational self-determination as it serves to protect the personal life and privacy of a people who entrusts themselves to certain professional groups.[27] Secrecy obligations apply vis-à-vis everyone and as a regular practice cover all the information that was entrusted to the obliged party in his professional capacity or of what has otherwise become known to him. Legislators tend to protect the duty of confidentiality with strong means, e.g. Section 203 of the German Criminal Code with the threat of a fine or imprisonment for violation of private secrets.[28] In many cases medical confidentiality demands higher data security standards. A necessary step in data processing is data sharing both internally and outside the hospital. Collegial sharing within departments, such as team meetings and shift changes require internal data transmission. It can be assumed that the lawfulness of the transmission is based on either the data subject's consent or on contractual necessity as far as the exchange of medical data is essential for the successful recovery. For data transmission to external stations, namely outsourcing, the consent of the data subject or another relevant legitimate basis is mandatory, especially when the purpose of the transmission deviates from fulfilling the treatment contract. Given the medical professional secrecy, outsourcing poses a high risk since the external service provider may acquire protected information.[29] The flexibility clause in Art. 90 (1) authorises Member States to introduce new rules regarding the powers of supervisory authorities vis-à-vis parties obliged to professional secrecy.

If additional provisions in addition to Art. 9 (1) exist, these must also be respected.[30] Such additional provision may be legally binding, such as Art. 28 (3) of the Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials of medicinal products for human use[31] or the Oviedo Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine, which is the only international legally binding instrument on the protection of human rights in the biomedical field applicable to the twenty-nine states that signed and ratified the Convention.[32] Although legally binding, however only applicable to certain circle of professionals is the Declaration of Helsinki on ethical principles for medical research involving human subjects.[33] Non-binding recommendations are released by international organisations[34] or by Member States professional and advisory bodies on a national level.[35]

[27] Reinhard Dettmeyer, *Medizin&Recht für Ärzte, Grundlagen – Fallbeispiele, Medizinrechtliche Fragen* (Springer, 2001) ch 4 73-74

[28] Ibid, 74-75; section 203 of the German Criminal Code (Strafgesetzbuch)

[29] Thomas Jäschke, *Datenschutz und Informationssicherheit im Gesundheitswesen* (ch 2, 2nd edn, MWV 2018) 69

[30] Weichert, Art. 9 para. 108, in Kühling and Buchner (eds) (n 21)

[31] Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials of medicinal products for human use, and repealing Directive 2001/20/EC [2014] L 158/1

[32] Council of Europe Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine (Oviedo Convention) [1997] ETS No. 164

[33] Word Medical Organisation Declaration of Helsinki adopted by the 18th WMA General Assembly, Helsinki, Finland, June 1964, last amended by the 64th WMA General Assembly, Fortaleza, Brazil, October 2013 [2013]

[34] eg. OECD Recommendations of the Council on Health Data Governance [2017] OECD/LEGAL/0433; OECD Interim Synthesis Report of October 2014 Data Driven Innovation for Growth and Well-Being, Chapter 8 The Evolution of Healthcare in a Data-Rich Environment [2015]

[35] for an overview of legally binding and non-binding instruments related to biomedical research and research with health data see Ciara Staunton and others, 'The GDPR and the research exemption: considerations on the

2.1.3. Processing concerning vulnerable individuals

R 75 lists processing activities where personal data of vulnerable natural persons are included as a risk factor to the rights and freedoms of those persons. This is a risk factor for the reason that the circumstances often prevent these individuals from understanding the implications of the processing of their data or how this could affect them. They might be unable to easily consent or oppose.[36]

This group may include children, elderly people, patients, mentally ill persons or dementia patients and any other case where there is an obvious imbalance between the data subject and the data controller.[37]

2.1.4. Large scale processing of personal data

Although the GDPR does not clarify what large-scale, 'extensive' processing means, some guidance can be found in R 91. According to factors mentioned here and by the Art. 29 WP large scale could refer to geographically widespread processing or processing involving a large number of data subjects and/or massive amount of personal data. Specific example for large scale is for example processing of patient data during a regular day in a hospital or by an insurance company as well as processing of personal data for behavioural advertising by a search engine.[38]

Given these criteria, it can hardly be questioned that Big Data analytics fall under this category. Big Data is usually described with the 4V (volume-variety-velocity-value) definition.[39] The sheer volume of the databases Big Data analytics usually target is by itself sufficient for these kinds of processing activities to be considered as large-scale processing.[40]

necessary safeguards for research biobanks' [2019] European Journal of Human Genetics 27, 1159, available at https://www.nature.com/articles/s41431-019-0386-5 accessed 04 December 2019

[36] ICO, 'Guidance on Data Protection Impact Assessments', available at https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/data-protection-impact-assessments-dpias/when-do-we-need-to-do-a-dpia/
accessed 26 September 2019

[37] Art. 29 WP, WP 218 (n 15) 10

[38] Art. 29 WP Guidelines on Data Protection Officer 16/EN WP 243rev.01, 21, available at https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612048 accessed 28 November 2019

[39] Sebastian Schulz, Art. 6 para. 254 in Sebastian Gola (ed) *Datenschutz-Grundverordnung: DSGVO Kommentar* (2nd edn C.H.Beck 2018) lately also 5V, volume-variety-velocity-veracity-value, see Anil Jain, The 5V's of big data, post for IBM, 17 September 2016 available at https://www.ibm.com/blogs/watson-health/the-5-vs-of-big-data/ accessed 26 September 2019

[40] Norbert Nolte and Christoph Werkmeister, Art. 35 para. 21 in Gola (ed) (n 39); Baumgartner, Art. 35 para. 23 in Eugen Ehmann and Martin Selmayr (eds), *Datenschutz-Grundverordnung Kommentar* (2 edn. C.H.Beck 2018); for more details on Big Data and a proposed concept of 'Big Personal Data' see Moira Paterson and Maeve McDonagh, 'Data Protection in an Era of Big Data: The Challenges Posed by Big Personal Data' Monash University Law Review (2018) 44:1-32, 2ff

2.1.5. Processing involving innovative use and newly introduced technological solutions

Art. 35 (1) and R 89, 91 stipulate that the use of new technology, defined "*in accordance with the achieved state of technological knowledge*" creates an obligation to carry out a data protection impact assessment. This is because application of such technologies imposes high risks to the individuals, especially when this is combined with additional risk factors.[41] Technological solutions on the cutting edge generally involve novel ways of data acquisition, sharing or usage. The risks lie in the novelty, best-practices and developed frameworks for the implementation may not be in place yet, or unknown societal consequences might surface.

Albeit the GDPR does not elaborate on what constitutes "state of the art", it implies that – within its provisions – these are technologies already made available to the public and are applied by practice and consequently, proved to be appropriate and suitable. It does not include newly introduced technologies or technologies only being tested.[42]

2.1.6. Analysis of combined and shared datasets

This risk factor gained relevance with the widespread utilisation of Big Data. The sense and purpose of Big Data is information extraction through fusion and evaluation of data from various origins.[43]

One consequence especially relevant from a data protection standpoint is that merging datasets might convert factual data otherwise with no reference to a specific person into data holding personally identifiable information.[44]

One further concern is that such processing might exceed the reasonable expectations of the data subjects, not to mention the constantly emerging new purposes of processing based on new findings. Since the search for correlations is often carried out generally without a specific target the possible results could be difficult to anticipate. This uncertainty is arguably challenging considering the principles of data processing regulated in Art. 5. In practice, controllers participating in a Big Data analysis should pay close attention to the obligations related to the transparency principle, Art. 5 (1) lit. a). It can be difficult to provide precise information about a processing, where data is obtained from several different sources, and these small inputs are later aggregated to produce a larger dataset.[45] Another principle of utmost importance in that regard is the principle of purpose limitation. Art. 5 (1) lit b). For the sake of better control, the purpose of the

[41] CIPL, Risk, High Risk, Risk Assessments and Data Protection Impact Assessments under the GDPR (White Paper, Cm, 2016) 9 available at www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_gdpr_project_risk_white_paper_21_december_2016.pdf accessed 26 September 2019

[42] Carlo Piltz, Art. 32 paras. 15, 16 in Gola (ed) (n 39); Mark Siebel, 'Abgrenzung der „allgemein anerkannten Regeln der Technik" vom „Stand der Technik' [2013] NJW 3000, 3003; cf. ECJ, decision of 29/5/1997 – C-300/95 *Commission of the European Communities v. United Kingdom of Great Britain and Northern Ireland* para. 26.

[43] Ira S. Rubinstein 'Big Data: The End of Privacy or a New Beginning?' (2013) Vol 3/ 2 IDPL 74, 76

[44] Christopher Werkmeister and Elena Brandt, 'Datenschutzrechtliche Herausforderungen für Big Data' [2016] CR 233, 234; Norbert Nolte and Christopher Werkmeister, Art. 35 para. 21 in Gola (ed) (n 39)

[45] Neil M. Richards and Johnathan H. King, 'Three Paradoxes of Big Data' (2016) Vol 66 Stanford Law Review 41, 42-43

processing shall be specified, explicit, legitimate and must be determined prior to the processing.[46]

2.2. Use case-related key challenges

This section discusses those data protection and data security issues directly linked to the use cases. Every issue can be associated with at least one of the identified risk factors in section 2.1.

The first is the question of the concept of personal data and anonymisation. It has been a year since the GDPR has come into force and it remains uncertain, where the line is between anonymous and personal data, what the threshold for anonymity is and how this can be realized. In addition, it has not been clarified, how exactly anonymisation affects the applicability of data protection provisions and, therefore, indirectly the obligations of the involved parties and the rights of the natural persons. The second and third challenges examine the conditions governing the lawfulness of processing and the purpose limitation principle. The last issue discussed in this section is the outsourcing of data processing activities.

2.2.1. The concept of personal data and anonymisation

2.2.1.1. Dynamics of data attributes

The current European data protection regulatory regime is built upon a binary concept of personal data.[47] Like two sides of the same coin, data by nature is either personal or non-personal, there is no third ‘state’ or category for anonymised – once personal – data.[48] Although this approach is practical in order to avoid sort of a ‘legal limbo’ and uncertainties, when thinking of anonymisation, it is extremely difficult to fit the dynamic data processing in this ‘base-2’ system.

The reason for that is the relativity that lies within the information the data holds.[49] Depending on the circumstances of the processing and utilisation the information might be associated with different type and extent of knowledge, therefore, the same data might be personal in one situation and at a given time whereas factual in another.[50] Even if at the beginning of the processing activity the data had no personal relevance, it should be regularly checked whether the data they use is still non-personal.[51] Also, with reasonable effort data originally relating to things can be brought into a direct ratio with a natural person, consequently, it may end up as personal data as well.[52]

This becomes conspicuous if examined through the lifespan of the data.

[46] Tobias Herbst, Art. 5 para. 31 in Kühling and Buchner (eds) (n 21); Art. 29 WP Opinion 03/2013 on purpose limitation [2013] WP 203 15 available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf accessed 22 November 2019

[47] Moritz Karg, Art. 4 Nr. 1 para. 14 in Simitis, Hornung and Spiecker (eds) (n 9)

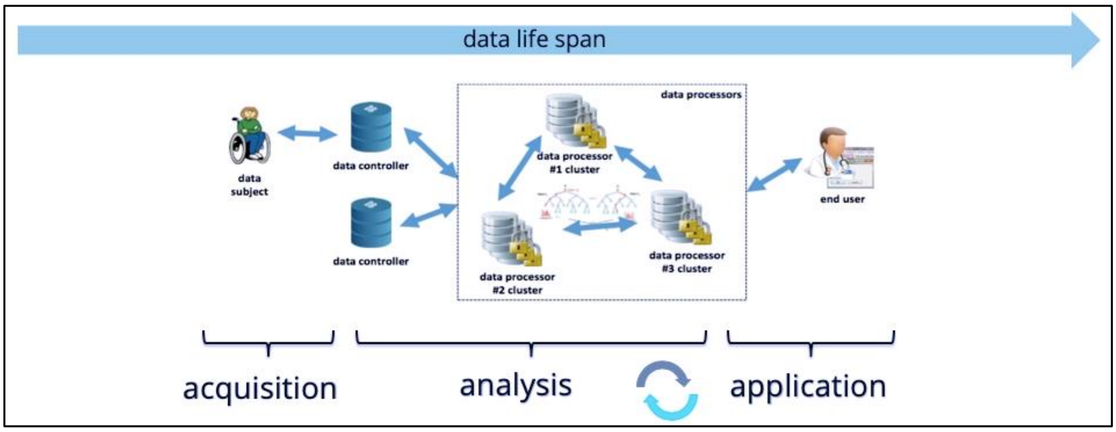
[48] Nikolaus Forgó, ’My health data —your research: some preliminary thoughts on different values in the General Data Protection Regulation’ (2015) International Data Privacy Law Vol. 5/1 54, 59

[49] information in this context is interpreted as data and the meaning associated with it as described in Mark Burgin, *Theory of Information. Fundamentality, Diversity and Unification* (World Scientific Publishing 2010) ch 2 181ff; for a detailed discussion on the significance of defining ‘information’ and ‘data’ in the modern legal and regulatory environment see Lee A. Bygrave, ‘Information Concepts in Law: Generic Dreams and Definitional Daylight’ (2015) Vol. 35 No. 1 Oxford Journal of Legal Studies 91-120

[50] within the context of this document the term “factual data” is used as a synonym for non-personal data

[51] Gerald Spindler, report, ‘Persönlichkeitsschutz im Internet - Anforderungen und Grenzen einer Regulierung’ in *Verhandlungen des 69. Deutschen Juristentages* (2012) Vol I F 116

[52] ibid, F 116



1. Figure Data life cycle in a SODA setting

Figure 1 shows a schematic representation of the three phases of the data life cycle within the frames of the SODA studies. Necessarily, the data life span starts by obtaining the data, at which stage the data is typically personal.[53] This is certainly the case within the frames of the SODA studies, where personal data is collected directly from the patients. After the acquisition the data is prepared for the analysis with a method called secret sharing and is divided into unintelligible shards. These modified data pieces are the input for the multi-party computation during the analysis. After the analysis the output data is protected with differential privacy, a statistical disclosure control technique.[54] This output data can be subject to various applications, a further analysis being one of them. While in the first stage the personal identifiability is an indivisible part of the data, the methods used further on aim to unlink the data from the natural person.
It can generally be said, that the nature of the data varies in the different stages of the Big Data processing and value chain. Being personal or factual is a dynamic rather than a static attribute of the data.[55]

2.2.1.2. The concept of personal data

Following this logic, the question naturally arises as to when it can be said that personal data has indeed became factual data.[56] Personal data ceases to be personal when it is converted into anonymised data, i.e. completely stripped from the natural person. Despite its tremendous practical relevance, the GDPR does not specify what it understands under anonymous data. It is only mentioned among the recitals as information the principles of data protection do not apply for.[57] It is presented as the inverse of personal

[53] Art. 29 WP Opinion 04/2007 on the concept of personal data [2007] WP 136, 2 available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf accessed 22 November 2019

[54] Alexandra Wood and others, 'Differential Privacy: A Primer for a Non-technical Audience' (2018) Vanderbilt Journal of Entertainment & Technology Law 21 No. 1, 209-275

[55] BJ Koops, 'The trouble with European Data Protection Law (2014) 4 (4) International Data Privacy Law 250, 252

[56] or, if it is even feasible; critical to the distinction between personal and non-personal data in the age of Big Data Paul Ohm, 'Broken promises of privacy: Responding to the Surprising Failure of Anonymisation' (2010) Vol. 57 UCLA Law Review 1701-1777

[57] R 26 s 5, 6 GDPR; note however, that albeit the GDPR does not apply for the processing of personal data, other regulations shall be considered, such as the Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union, OJ L 303/59

data, as data with no personal relevance whatsoever.[58] Conversely, this means that the broader the concept of personal data is, the higher the standards for anonymisation get. In other words, following an extensive understanding of personal data makes it more difficult to reach true anonymity in a legal sense. The concern of disproportionate extension of data protection rules was expressed by AG Jääskinen in his opinion in the recent Google Spain case. He argued that "*the broad definition of personal data* […] *is likely to cover an unprecedently wide range of new factual situations*" and advised the Court to take into account "*the principle of proportionality* […] *in order to avoid unreasonable and excessive legal consequences"* and to "*strike a correct, reasonable and proportionate balance*" between data protection and economic interest.[59] The Court did not follow the reasoning of AG Jääskinen but emphasised that data protection and the right to privacy are fundamental rights and for the purpose of an effective protection the proportionality shall be taken into consideration only when applying the data protection clauses, but not with the intention of circumventing or narrowing the scope itself.[60] The GDPR follows this rather broad understanding in defining its material scope and, ultimately, personal data.

Even though there is no abstract theoretical definition of anonymous data in the GDPR, positioning it as an opposite to personal data allows for an operational definition. In lights of this, it is necessary to examine the definition of personal data first.

Personal data is defined in Art. 4 Nr. 1 GDPR as

> „*any information relating to an identified or identifiable natural person ('data subject')*"

where an identifiable natural person is described as someone

> „*who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person*".

The four building blocks of this definition are:

1. "any information"	= attributes of the information
2. "relating to"	= relatability
3. "identified or identifiable"	= identifiability
4. "natural person"	= with regard to a natural person.

The first and the last elements do not need detailed discussion. "*Any*" in this context means practically anything that is available on a natural person, regardless the nature or format of the data.[61] Content-wise

[58] Manon Oostveen, 'Identifiability and the Applicability of Data Protection to Big Data' (2016) Vol. 6/4 International Data Privacy Law 299, 307

[59] Case C-131/12 *Google Spain SL/ Google Inc. v AEPD/Mario Costeja Gonzales* [2014] Opinion of AG Jääskinen para. 30-31

[60] ECJ decision of 13/05/2014 – C-131/12 *Google Spain SL/ Google Inc. v AEPD/Mario Costeja Gonzales* ECLI:EU:C:2014:317 para. 66-69; Nadezhda Putrova, 'The law of everything. Broad concept of personal data and future of EU data protection law' (2018) Innovation and Technology, 10:1, 40, 62

[61] Art. 29 WP, WP 136 (n 53) 6-7

it encompasses objective as well as subjective opinions and assessments.[62] As to the fourth element, the GDPR applies neither to data of legal entities nor to data of the deceased.[63]

For the purposes of this project much more interesting are the second and the third elements. Both relatability and identifiability are threshold conditions determining whether the data falls within the meaning of personal data. Therefore, the fact whether personal data can be thought of as anonymised depends on whether the conditions 'related to' a natural person and 'identified or identifiable' natural person are met or not.

As the definition in Art. 4 Nr. 1 GDPR stipulates, data must be relating to the data subject in order to be considered as personal data. This can be displayed as a content element, purpose element, or result element.[64] As regards to content the relation might exist directly or indirectly. All clear statements about an individual have direct relation to them, e.g. "*XY was brought to the hospital with symptoms indicating severe diabetes*". However, in certain situations data about an object can be informative about a certain individual. For example, a statement about the cost of an expensive medical device is formally related to the object, but at the same time implicitly contains information about the user of the device.[65] The second, purpose element is usually the reason for an information to be relating to an individual in case of profiling, where more information is combined in order to evaluate specific persons or treat them in a certain way.[66] In this case, as a result of the accumulation of different data from different sources a concrete individual is to be isolated by means of recognition of patterns within unstructured datasets. The third way data can be associated with a natural person is by result. A result or impact element is present when the processing of the data is likely to have either positive effects or repercussions on an individual's situation, rights or interests.[67] Accordingly, it is not necessary that the data as such focuses on a specified individual, it is enough that the outcome of the data processing somehow affects them.[68] This three constituent elements are not cumulative conditions, but alternative ones, that is where one element is fulfilled, there is no need for the other two to be present simultaneously.[69]

When examining the question of relatability from the SODA perspective, the reason for data to be associated with a specific person is at first sight the content of the data. Medical data is by nature highly sensitive personal data. Insurance claim data, which is analysed in the third pilot case, does not by default, fall under the special categories of personal data according to Art. 9 (1) GDPR, nonetheless, it could very well relate to an individual. As to the purpose element, creating a profile and targeting a specified individual is not the purpose of the SODA project. On the contrary, SODA does not develop tools for a targeted analysis, but for joint analysis of data from several sources for the purpose of non-personalised pattern recognition.[70] Where aggregated and statistical data cannot be associated with a person but with a group of persons, such data shall not relate to a person within the meaning of this constituent element.[71] However, the scope of this argument shall be interpreted strictly referring only to the purpose element.

[62] ECJ decision of 20/12/2017 C-434/16 *Peter Nowak v Data Protection Commissioner*, ECLI:EU:C:2017:994 para. 34

[63] R 27, 158, 160 GDPR; Moritz Karg, Art. 4 Nr. 1 para. 38-43, in Simitis, Hornung and Spiecker (eds), (n 9)

[64] Art. 29 WP, WP 136 (n 53) 9ff

[65] cf. example in Achim Klabunde, Art. 4 para.10 in Ehmann and Selmayr (eds) (n 40)

[66] Art. 29 WP, WP 136 (n 53) 10

[67] Manuel Klar and Jürgen Kühling Art. 4 Nr. 1 para. 14 in Kühling and Buchner (eds) (n 21)

[68] Art. 29 WP, WP 136 (n 53) 12

[69] ECJ, decision of 20/12/2017 – C-434/16 *Peter Nowak v. Data Protection Commissioner* ECLI:EU:C:2017:994 para. 35 (n 62); Art. 29 WP, WP 136 (n 53) 11

[70] for a detailed explanation of the distinction between categories of Big Data analytics in a legal sense see Sebastian Schulz, Art. 6 para. 256ff. in Gola (ed) (n 39)

[71] Manuel Klar and Jürgen Kühling Art. 4 Nr. 1 para. 15 in Kühling and Buchner (eds) (n 21); Sebastian Gola, Art. 4 Nr. 1 para. 8 in Gola (ed) (n 39) argues the same with regard to collective information on a given group of people, if the individual is not marked as a member of that respective group.

Last but not least the definition in Art. 4 Nr. 1 is based on the presence of personal reference and requires the identification or identifiability of a natural person. Undoubtedly, this part of the provision is the most discussed and most controversial aspect of the definition.

When a natural person is identified, is not further discussed. This first alternative refers to the act of identification as a past action, where the individual is or has already been distinguished from others. Generally speaking, someone is identified when he or she is specifically detected in a group of individuals.[72] The most obvious form is the so-called handshake identification, namely, when the data collection takes place in the presence of the given person.[73] As mentioned above, it has to be noted that while the same information may identify someone in one case, it does not in another, since the contextual element always has to be considered. The identification might happen directly with help of direct identifiers or indirectly. The second alternative concerns an indirectly identifiable natural person. A natural person is identifiable, when there is a possibility of identification, but it has not happened yet.

Basically, during the course of the applicability of the Data Protection Directive,[74] two opposing positions have been developed by data protection supervisory bodies and in legal literature to this question, the so-called absolute concept and the relative concept of identifiability. While the first one is more objective and rather inflexible as regards to determining the threshold for identifiability, the latter is a more context-sensitive approach.

According to the **absolute concept of identifiability**, information is regarded as personal data if the data controller or any other third party is able to connect the information to a specific person. It takes into account all means and possibilities available for the data controller or any third party. The theory is called absolute because it advocates for an unconditional applicability of data protection rules in a way that even a theoretical possibility of identification, regardless how unlikely, invokes legal effects.[75] This approach does not leave any kind of leeway and does not consider the principle of proportionality, since the individual abilities and means of the data controller or third parties remain disregarded. It is based on an extensive interpretation of the definition of personal data and handles the exceptions rather restrictively. In other words, for the establishment of personal reference it remains irrelevant whether data controllers or third parties make use of existing possibilities of linking information to an individual.[76] Additionally, a personal reference shall be considered present in case the linkage is technologically possible, yet prohibited by law.[77] In essence this theory implies that the potential existence of someone somewhere, who holds the additional knowledge that could link data to specific person is sufficient enough to consider a data subject identifiable.

Moreover, given that this approach does not deem a strong and explicit legal prohibition – or, conversely, a strong contractual protection – sufficient enough to eliminate the chance of identifiability, unlawful ways of data access, such as unauthorised access or hacking shall be taken into account as well. This leads to a hardly feasible and often impossible situation that the only case personal data could ever lose its identifiable nature is when the original raw data is deleted and does not in any form exist anymore. Precisely in case of data processing by hospitals or insurance companies the statutory storage obligations and reasonable retention periods do not allow for a deletion of the data at the time the research starts provided that the purpose of the original processing has not yet been achieved, or there are parallel running archiving duties and periods.

[72] Art. 29 WP, WP 136 (n 53) 12

[73] Moritz Karg, Art. 4 Nr. 1 para. 55 in Simitis, Hornung and Spiecker (eds) (n 9)

[74] Directive 1995/46/EC of the European Parliament and of Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movements of such data, OJ L 281/31.The Directive was repealed with the effect from 25 May 2018 in accordance with Art. 94 (1) GDPR.

[75] Stefan Brink and Jens Eckhardt, ’Wann ist ein Datum personenbezogenes Datum? – Anwendungsbereich des Datenschutzes’ [2015] ZD 205, 206

[76] Moritz Karg, Art. 4 Nr. 1 para. 58 in Simitis, Hornung and Spiecker (eds) (n 9)

[77] ibid para. 58

Following this logic, anyone wishing to design data processing activities in such a way that with the application of privacy-preserving tools they would avoid the pitfalls of data protection law shall be prepared to provide a "negative evidence" and demonstrate that the assignment of data to a person can indeed be completely, *absolutely* ruled out.[78] Consequently, if the pure form of the absolute theory prevailed, there would be virtually no real playing field for anonymous data at all, since it is, in fact, not possible to prove that there is no one who could have access to the information and to the means needed to link it to an individual.[79]

It can be said regarding privacy preserving methods, as long as there is any chance, that anyone in the world is able to get access to the raw dataset, plaintext or decipher an encrypted dataset, the processing of the data would always be subject to data protection legislation, regardless the used technique. For Big Data analytics using encryption or cryptography this would mean that regardless of the implemented technical measure or combination of technical measures the essential character of personal data would not change, only the personal data would be better protected against unauthorised access by intruders. However, the obstruction of access does not mean the impossibility of access. In case of transferring raw data to a storage provider using state-of-the-art encryption for example, the storage provider would be a data processor, even though he is never in possession of the decryption key. The data controller as transferring party holds the decryption key – the additional knowledge – that in theory would allow the identification and the existence of this low-probability re-identification excludes the rejection of personal reference. Therefore, the receiving party would be data processor and subject to obligations of data protection law.[80] An equally relevant example is the transferring of anonymised data.

The other, opposing theory is the subjective or **relative concept of identifiability**, which is based on the assessment of realistic chances of the data controller used to identify the individual. The grave difference from the abovementioned absolute theory is that according to the relative approach only those means are to be taken into account that could actually be applied by the respective data controller in the concrete individual case in order to establish personal reference. Relevant factors in determining whether a data controller has the means, opportunity and knowledge must be considered, a merely hypothetical event remains out of consideration.[81] Furthermore, according to this theory, personal reference would be excluded if the only way it could be (re)produced was by using illegal means.[82]

It accepts the existence of certain low risks and at the same time leaves those low-probability events of potential (re)identification out of the equation when assessing the absence of identifiability.[83] Essentially, this approach emphasises the plausibility of and rationale behind the actions of the data controller and third parties bearing in mind economic and legal feasibility.

This context-sensitivity makes the relative approach much more favourable for Big Data analytics using privacy-preserving tools. Taking encryption for example, in situations where the data controller has or

[78] Niko Härting, para. 267 in Niko Härting, *Datenschutz-Grundverordnung* (Otto Schmidt Verlag, 2016)

[79] this view is consistent with the approach of the ISO standard on health informatics, where anonymisation is defined as „*process by which personal data […] is irreversibly altered in such a way that a data subject can no longer be identified directly or indirectly, either by the data controller alone or in collaboration with any other party*" and it is noted that this „*concept is absolute, and, in practice, it may be difficult to obtain.*" ISO 25237:2017 Standard on health informatics – Pseudonymisation [2017] Nr. 3 3.2 available at https://www.iso.org/obp/ui/#iso:std:63553:en accessed 05 December 2019

[80] Ninja Marnau, 'Anonymisierung, Pseudonymisierung und Transparenz für Big Data' (2016) 40 DuD 428, 430

[81] ibid para. 264; Gabriele Buchholtz and Rainer Stentzel, Art. 4 Nr. 1 para. 8 in Sibylle Gierschmann and others (eds) *Kommentar Datenschutz-Grundverordnung* (Bundesanzeiger Verlag, 2018); Alexander Roßnagel and Philip Scholz, 'Datenschutz durch Anobymität und Pseudonymität – Rechtsfolgen der Verwendung anonymer und pseudonymer Daten' [2000] MMR 721, 723; Michael Knopp, 'Pseudonym – Grauzone zwischen Anonymisierung und Personenbezug' (2015) 39 DuD 527, 529

[82] Moritz Karg, Art. 4 Nr. 1 para. 59 in Simitis, Hornung and Spiecker (eds) (n 9)

[83] Samson Esayas, 'The Role of Anonymisation and Pseudonymisation Under the EU Data Privacy Rules: Beyond the 'All or nothing' Approach' (2015) 6 No. 2 European Journal of Law and Technology, 2 available at SSRN: https://ssrn.com/abstract=2746831 accessed 5 December 2019.

with reasonable expectations is able to obtain additional knowledge necessary for identification, e.g. decryption key in case of encryption, the encrypted data was personal data and thus data protection law would apply to the processing.[84] Generally speaking, in case of outsourced data analysis where the cloud user holds the decryption key or the additional knowledge in another form, but the data cannot be converted legible without this by the cloud provider, the data can be considered non-personal data in relation to the cloud provider. It must be noted, however, that the data does not cease to be personal for the party holding the necessary information. Taking the same example of transferring state-of-the-art encrypted data to a storage provider, according to the relative approach of identifiability the data could be seen anonymised for the storage provider, provided that he neither has access to the means of re-identification nor can bypass the system on a technical level.[85] Consequently, the storage provider would not be a data processor and his actions would not be governed by data protection rules.

The extent to which the knowledge and means of third parties and the situation-dependent circumstances are to be taken into account for identification remains controversial under the GDPR. A number of hybrid forms can be found between the two extreme positions and the GDPR did not take the chance and did not fully endorse neither the absolute nor the relative approach. A description is provided by R 26 GDPR:

> *„The principles of data protection should apply to any information concerning an identified or identifiable natural person. [...] 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. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments. The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable."*

The GDPR is rather ambiguous and is therefore subject to different interpretations. In an effort to unscramble the meaning and practical implications of the GDPR it is worth examining the opinion of the Art. 29 WP and the development of the ECJ's case law on the concept of personal data.[86]
The Art. 29 WP has dealt with the question of identifiability in its opinion on the concept of personal data.[87] Next to the quite clear case of direct identification the opinion discusses indirect identification in great detail. According to the opinion a mere hypothetical possibility of recombination is not sufficient yet there are "*unique combinations*", circumstances where the probability of identification is higher than under different circumstances. The benchmark criteria for the possibility of identification represented by the opinion are "*the means likely reasonably to be used by the controller or any other person".* Interestingly, these criteria include elements of both the relative and the absolute approach and combine them into an objective concept opened to context-sensitivity. Firstly, considering not only

[84] Gerald Spindler and Philip Schmechel, 'Personal Data and Encryption in the European General Data Protection Regulation' (2016) 2 JIPITEC available at https://www.jipitec.eu/issues/jipitec-7-2-2016/4440 accessed 01 December 2019

[85] Ninja Marnau, 'Anonymisierung, Pseudonymisierung und Transparenz für Big Data' (2016) 40 DuD 428, 430 (n 80)

[86] for a more detailed, in-depth discussion of the Art. 29 WP Opinion on the concept of personal data and the ECJ's Breyer decision in general see Gerald Spindler and Anna Zsófia Horváth, 'D3.1 General Legal Aspects' [2017] deliverable produced within the frames of the SODA project, available at https://www.soda-project.eu/wp-content/uploads/2018/02/SODA-D3.1-General-Legal-Aspects.pdf accessed 10 December 2019

[87] Art. 29 WP, WP 136 (n 53) 13ff

the controller but also any other person is a clear trait of the absolute theory. On the other hand, the WP explicitly restricts the circle of measures that needed to be taken into account and at the same time highlights the importance of contextual elements. Thus, even if the means by the controller and any other person are relevant and need to be assessed, these "means" are limited to those which are likely reasonably to be used in each specific case.[88]

The "*factors at stake*" needed to be evaluated during the assessment are:

1. costs of conducting the (re)identification
2. intended purpose of the processing
3. structure, design of the processing
4. advantage expected by the controller
5. interests at stake for the individual
6. risks of organisational dysfunctions, such as data breach or confidentiality issues
7. technical failures
8. state-of-the-art nature of applied technical and organisational measures at the time of the processing.[89]

The opinion accepts that unlawful means, such as a potential external hacking shall be taken into account in determining the likelihood of identifiability as well. Furthermore, it states that the implementation of appropriate technical and organisational measures is a prerequisite, which, if fulfilled ensures that the data is not personal data anymore.[90] Albeit the adoption of a list with the factors outlined above itself means a departure from an extreme absolute theory, the understanding of "*reasonably likely*" remains very broad.

The ECJ addressed the same issue in its recent Breyer-case.[91] The case was referred to the Court as a request for preliminary ruling procedure by the German Federal Court of Justice (BGH) for guidance on the interpretation of the dispute whether dynamic IP addresses can be considered as personal data.[92] After referring to the disagreement relating to the absolute or relative criterion, the German Federal Court of Justice opted for a relative approach in its request, and argued, that the consideration of the means a third party has to identify an individual does not necessary eliminate a relative approach, inasmuch as only the means realistically be used are taken into account.[93] In its judgement the ECJ opted for a two-fold reasoning, the two aspects being 1) all the means *reasonably likely* to be used shall be taken into account, 2) not only by the data controller, rather *by the data controller* or *by any other person*.[94] This two-level argument basically echoes the view of the Art. 29 WP in WP136.[95]

The novelty and significance of the case lies in the standpoint of the Court regarding legal prohibitions. In his opinion released on the case AG Sánchez-Bologna argued, that the systematic interpretation of

[88] ibid, 15, 21
[89] ibid, 15, 17
[90] ibid, 17
[91] ECJ decision of 19/10/2016 – C-582/14, *Patrick Breyer v Bundesrepublik Deutschland*, ECLI:EU:C:2016:779
[92] German Federal Court of Justice (BGH), decision of 28/10/2014 - VI ZR 135/13 = MMR 2015, 131; regarding the classification of IP addresses as personal data for access providers judged by the EJC decision of 24/11/2011 – Case C-70/10, *Scarlet Extended SA v Sabam* ECLI:EU:C:2011:771 para. 51 which states that *"[IP] addresses are protected personal data because they allow those users to be precisely identified*"
[93] Breyer (n 91) [21], [25]; German Federal Court of Justice (BGH), decision of 28/10/2014 - VI ZR 135/13 = MMR 2015, 131, 133f
[94] Breyer (n 91) [42]
[95] Putrova (n 60) 64

the Data Protection Directive[96] implies that means are not reasonably likely to be used when contacting the other party who holds additional information necessary for identification "*is, in fact, very costly in human and economic terms, or practically impossible or prohibited by law*".[97] The Court accepted this argument and it concluded in line with the AG's opinion that the means used for combining data from different sources is not reasonably likely to be used if "*the identification of the data subject was prohibited by law or practically impossible on account of the fact that it requires a disproportionate effort in terms of time, cost and man-power*", and thus "*the risk of identification appears in reality to be insignificant*".[98] Only if the linking is legally permissible *and* the access to the means and the knowledge of other parties is reasonably likely by the party concerned is there a personal reference.[99] In so far as the ECJ does not take the illegal means or any kind of disproportional effort into account, this argumentation is a step towards a relative approach.[100] Forming the scope of what constitutes unreasonable likelihood the Court adopted an interpretation that restricts, "relativizes" a pure absolute approach. It is a rather progressive concept that puts the given data processing activities into a context and emphasises the significance of technical and legal circumstances. [101]
In essence, both the Art. 29 WP and the Court represents an argument consistent with the absolute concept with relative, context-sensitive elements. The difference is that in the Breyer-decision the Court took a more sophisticated approach and added that the fact that combining data for the purpose of re-identification is prohibited by law would mean that the means of re-identification are "*less reasonably likely to be used*".[102]

Similar to R 26 of the Data Protection Directive, the opinion of the Art. 29 WP and the statements of the ECJ in the Breyer-decision R 26 of the GDPR underlines the importance of the means reasonably likely to be used as a standard for identifiability. Further, it states that not only the means by the data controller shall be considered but also the means of any other third party. R 26 offers a non-exhaustive list of factors that are decisive for the likelihood of means to be applied, listing "*costs of and the amount of time required for identification, taking into consideration the available technology at the time of the*

[96] R 26 of the Directive (n 74) applicable at the time „*[...]whereas, to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person*"

[97] Case C-582/14 *Patrick Breyer. v Bundesrepublik Deutschland* [2016] Opinion of AG Sánchez-Bologna para. 68

[98] Breyer (n 91) [46]; Klaus Brisch and Fritz Pieper, 'Das Kriterium der "Bestimmbarkeit" bei Big Data-Analyseverfahren' [2015] CR 724, 728, who argue that the word „reason" is indeed not consistent with the use of illegal means, however, is against for a complete exclusion of "the illegal means" and calls for a case-by case consideration; Alessandro El Khoury, 'Dynamic IP Addresses Can Be Personal Data, Sometimes. A Story of Binary Relations and Schrödinger's Cat' (2017) 1 European Journal of Risk Regulation 191, 195 is rather critical about what may be "*prohibited by law*", same conclusion in Denis Kelleher, 'In Breyer decision today, Europe's highest court rules on definition of personal data' (*Iapp Blog*, 19 October 2016) available at https://iapp.org/news/a/in-breyer-decision-today-europes-highest-court-rules-on-definition-of-personal-data/ accessed 10 December 2019

[99] Moritz Karg, Art. 4 Nr. 1 para. 62 in Simitis, Hornung and Spiecker (eds) (n 9)

[100] El Khoury (n 98) 196, supports the view of "*double relativity*", based on the phrase "*means likely reasonably to be used*" of the Data Protection Directive as well as the phrase "*disproportional*" in the ECJ's judgement

[101] Manuel Klar and Jürgen Kühling, Art. 4 Nr. 1 para. 28 in Kühling and Buchner (eds) (n 21); Non-disclosure agreements may be one way to guarantee the impediment of identification, since such contractual clauses are legal prohibitions of recombining data from different providers. Note, that a non-disclosure agreement does not mean data processing agreement between the controller and the processor according to Art.28 (3) GDPR, it is merely an agreement by which the parties are bound not to disclose certain information, Heidi Waen, Jaqueline Van Essen and Vincent Wellens, 'Confidentiality agreements are not data processing agreements' (Lexology, 8 September 2015) available at https://www.lexology.com/library/detail.aspx?g=e1d5ccfb-f0a0-4c32-aa59-a65864af1acd accessed 10 December 2019

[102] Putrova (n 60) 64-65 comes to the same conclusion

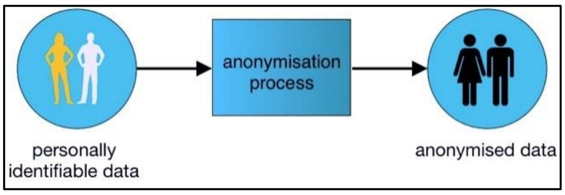
processing and technological developments". These factors set out the frames of the proportionality established by the phrase "reasonably likely" and thus they point towards a relative concept.[103] Significant factors are the general expenses of the party motivated to obtain the data and to identify the data subject, as well as the state of science and technology, including foreseeable future technical development.[104] The word "objective" as provided in Recital 26 means that the relevance of the factors in each case is decided by general economic factors as unbiased standard, not by discretion of the involved controller or the another person.[105] The wording of the new definition of personal data in Art. 4 No. 1 GDPR as well as the matching Recital 26 are very similar to the old ones in the Directive. Hence it can be assumed that the interpretation outlined by the ECJ judgement in the Breyer-case continues to apply.

As it can be seen from the wording of R 26 the GDPR refuses a categorically one-sided relative concept and follows the layered approach previously adopted by the Art. 29 WP and the ECJ. Albeit the high requirements the GDPR imposes for the assessments of identifiability, data could indeed be considered non-personal data for the parties who do not possess auxiliary information needed for identification provided that there are no means reasonably likely to be used by them to obtain such information.[106] These findings support the notion that identifiability under the GDPR shall be answered based on a test that in itself follows a relative concept, but includes the assessment of objective criteria.

2.2.1.3. Anonymous data and a two-phase model of anonymity

The previous analysis of the elements of the definition of personal data according to Art. 4 Nr. 1 GDPR has shown that the GDPR adopted a rather extensive understanding of personal data. As stated above, anonymous data is presented in the GDPR as the inverse of personal data. Therefore, the broader the scope of personal data, the narrower the leeway for anonymous data. From the perspective of anonymous and anonymised data, the most influential element of the personal data definition in Art. 4 Nr. 1 is undoubtedly the indirect identifiability. An effective anonymisation depends on the understanding of what constitutes identifiable personal data. Against this backdrop it would appear that personal data could be rendered anonymous if there is no chance reasonably likely to be used for re-establishing the personal reference.

A dual concept of anonymisation could offer a solution on a conceptual level. This model handles anonymisation as processing of the data separately from anonymity as a 'state' of the data, as shown on Figure 2.



2. Figure Two-phase model of anonymity: 1) anonymisation as 'processing' of personal data under Art. 4 Nr. 2 GDPR and 2) anonymised data as outcome, as a 'state' of the data

According to the opinion advocated here, the anonymisation process constitutes a processing of personal data, irrespective of the used technique. It falls under the material scope of the GDPR and all

[103] Gerald Spindler, 'Die neue EU-Datenschutz-Grundverordnung' (2016) Der Betrieb 937, 937ff

[104] Härting (n 78) [284]; Stefan Ernst, Art. 4 paras. 10, 11 in Boris Paal and Daniel Pauly (eds) *Datenschutz-Grundverordnung* (2nd edn. C.H. Beck 2018)

[105] Marnau (n 80) 430

[106] ibid, 430; Brink and Eckhard (n 75) 209

the principles of processing and obligations of the data controller and processor relating to the processing of personal data apply.[107] However, interestingly, whether anonymisation as such constitutes a further processing of personal data turns out to be a quite controversial issue in the legal literature.[108] In order to shed some light on the validity of this question, it is advisable to examine the definition of processing first. Art. 4 Nr. 2 GDPR offers a definition for processing of personal data as follows:

> "*processing' means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction; [...]*"

It does not include the phrase anonymisation, or any technical requirements that could imply anonymisation. This absence of term has given rise to theories arguing that anonymisation cannot be subsumed under the term processing if the specific method used for anonymisation cannot be matched with one of the terms on the list. The fact that the GDR relies on more general-abstract terms comes hardly as a surprise, since any mentioning of specific techniques would contradict the technological neutrality of the regulation. As R 15 outlines *"the protection of natural persons should be technologically neutral and should not depend on the techniques used"*.

One way to overcome this problem is to think of anonymisation as a generic term which labels a wide range of de-identification techniques. It includes various different methods and is therefore a heterogeneous category of technological solutions. Accordingly, a given anonymisation process can be implemented through alteration, re-structuring or by partial deletion, that is through aggregation by removing attributes from the dataset.[109] Following this logic, albeit the term anonymisation is not listed in Art. 4 Nr. 2, every step of a given anonymisation process could be equated with a processing operation listed in Art. 4 Nr. 2. In other words, accepting the view that each technical operationalisation has an equivalent, then all anonymisation really is, is the combined implementation of several processing operations named by Art. 4 Nr. 2. Therefore, assuming that the listed processing operations are sub-categories, '*building blocks*' in relation to anonymisation as such, one could argue that the GDPR regulates anonymisation through regulating the basic processing operations it is ultimately built

[107] Data Protection Commission of Ireland, 'Guidance Note: Guidance on Anonymisation and Pseudonymisation' (2019) 13 available at https://www.dataprotection.ie/sites/default/files/uploads/2019-06/190614%20Anonymisation%20and%20Pseudonymisation.pdf accessed 22 November 2019; ICO, 'What about anonymised data?' available at https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/what-is-personal-data/#pd5 accessed 22 November 2019; this view is in accordance with the standpoint adopted by the Art. 29 WP as well, Art. 29 WP Opinion 05/2014 on anonymisation techniques [2014] WP 216, 3, 7, available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf accessed 22 November 2019

[108] a number of authors have accepted anonymisation as processing of personal data Alexander Roßnagel, Art. 4 Nr. 2 paras. 14, 26 in Simitis, Hornung and Spiecker (eds) (n 9); Wolfgang Ziebarth Art. 4 para. 28 in Gernot Sydow (ed) *Europäische Datenschutz-Grundverordnung Handkommentar* (Nomos, 2017); same conclusion in Kai-Uwe Plath, Art. 4 para. 9 in Kai-Uwe Plath (ed) *Kommentar zu DSGVO, BDSG und den Datenschutzbestimmungen von TMG und TKG* (3th edn Verlag Dr Otto Schmidt KG 2018); Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V. Working Paper 'Arbeitshilfe zur Pseudonymisierung /Anonymisierung' (2018) 9, 10 available at https://www.gesundheitsdatenschutz.org/download/Pseudonymisierung-Anonymisierung.pdf accessed 22 November 2019; whereas against this view argue in relation to adaptation and alteration Tobias Herbst, Art. 4 para. 25 in Kühling and Buchner (eds) (n 21); Stefan Ernst, Art. 4 para. 27 in in Boris Paal and Daniel Pauly (eds) *Datenschutz-Grundverordnung* (2nd edn. C.H. Beck 2018) (n 104)

[109] note however, that anonymisation does not necessarily mean deletion or vice versa, Alexander Roßnagel, Art. 4 Nr. 2 paras. 31-32 in Simitis, Hornung Spiecker (eds) (n 9)

up from. Consequently, a new anonymisation tool can be developed by the arbitrary combination of those single operations.

This theory is based on the rather extensive interpretation of "processing" the GDPR follows.[110] Any operation that directly or indirectly affects personal data as well as any form of treatment of personal data represents a "processing" activity in the sense of the GDPR.[111] 7 It was an on-purpose decision that the GDPR does not contain any further clarification of what "automated means" of processing are. As argued above, from the background of rapid technological developments it was necessary in order to stay technologically neutral.[112] In this way no anonymisation method is preferred or privileged.

This analysis presents convincing evidence that anonymisation shall be considered as data processing in accordance with Art. 4 Nr. 2 GDPR. In light of the abovementioned arguments three main findings emerge:

1. on the input level: at the beginning, that is during the early stage of the processing and during the collection certainly processing of personal data, since the aim of the following anonymisation is to separate the data from obvious identifying attributes
2. during the process of de-identification: "grey", since the GDPR only knows a black-and-white approach it needs to be assumed that the data is rather personal data. In this way there is no protection gap in relation to the data subject in case of a data breach.
3. on the output level: non-personal data, provided the anonymisation process is successful.

Presuming the anonymisation is executed as planned and with the reservation that there are no means reasonably likely to be used available for re-identification, the output data can be considered anonymised. In this case, the further handlings of this anonymous data fall outside the scope of the GDPR and the provisions of data protection do not apply anymore.

The question as to under which circumstances is an anonymisation process successful has attracted more attention from the Art. 29 WP in their opinion on anonymisation techniques.[113] Following the interpretation of this opinion anonymisation given the current state of technology must work as permanent as erasure, i.e. it has to be able to make it impossible to process personal data.[114] At first glance this concept is closely related to the abovementioned absolute approach. This would indeed be a case, if the view of the Art. 29 WP on anonymisation weren't compatible with the a more flexible, relative concept. In order to determine which anonymization methods are suitable for eliminating the presence of and access to means reasonably likely to be used for re-identification and are, therefore, most likely able to "reliably produce anonymised information", the Art.-29 Working Party provided a guide on efficiency of the most common anonymization techniques.[115] According to this analysis the anonymisation must be robust against three types of identification threats:

1. **Singling out**: possibility to isolate records of an individual in the dataset
2. **Linkability**: ability to link, at least, two records concerning the same data subject or group of data subjects in the same database or in two different data bases

[110] Jürgen Kühling and Johannes Raab, Art. 2 para. 13 in Kühling and Buchner (eds) (n 21)
[111] Kai-Uwe Plath, Art. 4 para. 9 in Plath (ed) (n 108)
[112] Jürgen Kühling and Johannes Raab, Art. 2 para. 15 in Kühling and Buchner (eds) (n 21)
[113] Art. 29 WP Opinion WP 216 (n 107)
[114] ibid, 6
[115] ibid, 11ff; for a technical perspective on risks related to anonymisation see Khaled El Emam, Sam Rodgers and Bradley Malin, 'Anonymising and sharing individual patient data' (2015) BMJ 350 available https://www.bmj.com/content/350/bmj.h1139 at accessed 21 November 2019

3. **Inference**: the possibility to deduce, with significant probability, the value of an attribute from the values of a set of other attributes.[116]

Apart from inference, re-identification and combining datasets which allow the emergence of patterns related to a single individual or a specific group another concern emerges from the sole understanding of Big Data analytics itself. Big Data Analytics mean the whole data management lifecycle, the acquisition, analysis and application of the data.[117]

It should be noted though that a certain anonymisation tool can serve as an "escape" from the regulations of the GDPR only as long as it is state-of-the-art and secure enough to prevent re-identification. The technological development might, in essence, influence whether data is personal or not just as much as the contextual elements of the processing activity. Therefore, on a practical level it is crucial to carry out a context-specific risk assessment as well as regular reviews and continuous evaluation accompanied by comprehensible documentation.[118]

2.2.2. Determination of legitimate basis of processing

2.2.2.1. Lawful basis of the processing of special categories of personal data

The principle of prohibition with the reservation of authorisation is the cornerstone of the processing of personal data under the GDPR.[119] According to this principle set out in Art. 6 (1) the processing of personal data is fundamentally prohibited unless it can be justified on the basis of the rules on the lawfulness of processing.[120] It demands that each and every processing activity carried out by the data controller and processor concerning personal data shall be based on a legitimate basis otherwise the entire processing is unlawful right from the outset. [121] It is therefore necessary, and also sufficient, to designate one of the substantially equivalent legitimate bases.[122] One processing activity cannot be based on multiple lawful bases. Data controllers must identify the appropriate lawful basis in advance and are not allowed to swap between lawful bases.[123] Art. 6 (1) GDPR names six legal bases, with the list being definitive and exhaustive. The six legal bases are:

1. consent to the processing for one or more specific purposes;
2. processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;
3. processing is necessary for compliance with a legal obligation to which the controller is subject;

[116] Art. 29-WP, WP 216 (n 107) 11, 12; see also: Stefan Ernst, Art. 4 para. 51 in Paal and Pauly (eds) (n 104); Mike Hintze and Khaled El Emam, *'Privacy Analytics White Paper – Comparing the Benefits of Pseudonymisation and Anonymisation Under the GDPR*' (*Iapp blog*, 17 August 2017) 11, available at https://iapp.org/media/pdf/resource_center/PA_WP2-Anonymous-pseudonymous-comparison.pdf accessed 21 November 2019

[117] ENISA, Working paper on 'Privacy and Data Protection by Design – From policy to engineering' (2015) 11 available at https://www.enisa.europa.eu/publications/privacy-and-data-protection-by-design accessed 12 November 2019; cf. Oostveen (n 58) 231ff, who calls this a „threephase-model" of Big Data

[118] Manuel Klar and Jürgen Kühling, Art. 4 Nr. 1 para. 22 in Kühling and Buchner (eds) (n 21)

[119] Benedikt Buchner and Thomas Petri, Art. 6 para. 11ff in Kühling and Buchner (eds) (n 21); Christoph Schulz, Art. 6 para. 2 in Gola, (n 39); Horst Heberlein Art. 6 para. 1, in Ehmann and Selmayr (eds) (n 40); Eike Michael Frenzel, Art. 6 para 1. in Paal and Pauly (eds) (n 104)

[120] R 40 GDPR

[121] cf. R 40 GDPR

[122] Gerald Spindler and Lukas Dalby, Art. 6 para. 2 in Gerald Spindler and Fabian Schuster (eds) *Recht der elektronischen Medien, Teil DS-GVO* (4th edn, C.H. Beck 2019)

[123] Art. 29 WP, Guidelines on Consent under Regulation 2016/679, WP 259, 22 available at https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051 accessed 20 November 2019

4. processing is necessary in order to protect the vital interests of the data subject or of another natural person;
5. processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;
6. processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

It is important to note at this point that the processing of special categories of data, such as health data is subject to additional protection. The processing of such data is allowed exceptionally and exclusively in case one of the conditions outlined in Art. 9 (2) applies. Pursuant to Art. 9 (2), the prohibition in para. 1 does not apply within the framework of the processing modalities specified in Art. 9 (2) lit. a-j). This reflects in Recital 51, which states that "*such personal data should not be processed, unless processing is allowed in specific cases set out in this Regulation, taking into account that Member States law may lay down specific provisions on data protection in order to adapt the application of the rules of this Regulation for compliance with a legal obligation or for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller. In addition to the specific requirements for such processing, the general principles and other rules of this Regulation should apply, in particular as regards the conditions for lawful processing*".[124]

Data controllers cannot in any case justify the processing of sensitive data by referring only to a lawful basis in Art. 6 (1), without taking into account Art. 9 (2). Especially evoking their legitimate interest pursuant to Art. 6 (1) lit f) is not compliant, since the Art. 9 (2) excludes the possibility of such a wide-ranging interest assessment. In addition to the specific requirements set out in paragraph 2, the general principles and other provisions of lawful processing of the Regulation (Art. 5-8) must however be heeded.[125] Consequently, for the lawful processing of sensitive data two consecutive requirements need to be fulfilled: 1) identifying a lawful basis under Art. 6 of the GDPR and 2) identifying a separate condition for processing under Art. 9.[126] Art. 9 (2) provides a definitive list, processing of special categories of personal data is only permitted under these special conditions:

1. explicit consent of the data subject;
2. necessary for carrying out of obligations under employment, social security or social protection law, or a collective agreement;
3. necessary to protect the vital interests of a data subject who is physically or legally incapable of giving consent;
4. the processing is carried out by a not-for-profit body with a political, philosophical, religious or trade union aim provided the processing relates only to members or former members (or those who have regular contact with it in connection with those purposes) and provided there is no disclosure to a third party without consent;
5. data manifestly made public by the data subject;

[124] R 51 s 4 GDPR

[125] R 51 s 5 GDPR

[126] ICO, 'Special category data' (*ICO Guide to the General Data Protection Regulation*) available at https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/special-category-data/ accessed 11 December 2019

6. necessary for the establishment, exercise or defence of legal claims or where courts are acting in their judicial capacity;
7. necessary for reasons of substantial public interest on the basis of Union or Member State law which is proportionate to the aim pursued and which contains appropriate safeguarding measures;
8. necessary for the purposes of preventative or occupational medicine, for assessing the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or management of health or social care systems and services on the basis of Union or Member State law or a contract with a health professional;
9. necessary for reasons of public interest in the area of public health, such as protecting against serious cross border threats to health or ensuring high standards of healthcare and of medicinal products or medical devices;
10. necessary for archiving purposes in the public interest, or scientific and historical research purposes or statistical purposes in accordance with Article 89 (1).

2.2.2.2. Use-case specific evaluation of relevant lawful bases

This section outlines those provisions of Art. 9 (2) that are particularly relevant within the context of the case studies.

2.2.2.2.1. Explicit consent – Art. 9 (2) lit a)

A valid consent for the processing of health data must fulfil a two-level set of requirements. Firstly, the general requirements for a valid consent continue to apply. The definition of consent within the frames of data protection rules is laid down by Art. 4 Nr. 11 GDPR as

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

Art. 4, however, only lists a non-exhaustive set of the conditions of a legally compliant consent. Only when the legal definition in Art. 4 is combined with Art. 6 (1) lit. a), Art. 7 and the special conditions in Art. 8 and 9 (2) lit. a) does a complete picture of the relevant requirements emerge.[127]

The core of the element "freely given" is the fact that the data subjects must have a genuine choice to accept or decline the terms of processing offered to them.[128] If there is any sign of compulsion, undue pressure or if negative consequences arise from the declination, the consent will not be valid. This is highlighted also in Recital 42 which clarifies that "*consent should not be regarded as freely given if the data subject has no genuine or free choice or is unable to refuse or withdraw consent without detriment*". Recital 43 addresses the situation where there is a clear imbalance between the data subject and the controller. This is often a case whenever a controller is a public authority, since in such cases the data subject does not have any realistic alternative to accepting the terms.[129] Additionally, Recital 43 states that consent "*is presumed not to be freely given if it does not allow separate consent to be given to different personal data processing operations despite it being appropriate in the individual*

[127] cf. Gerald Spindler and Lukas Dalby, Art. 7 para. 4 in Spindler and Schuster (eds) (n 122)

[128] Art. 29 WP, WP 259 (n 123) 4, 6; Katrin Schaar, 'Anpassung von Einwilligungserklärungen für wissenschaftliche Forschungsprojekte – die informierte Einwilligung nach der DS-GVO und den Ethikrichtlinien' (2017) ZD 213, 214

[129] Art. 29 WP, WP 259 (n 123) 7

case". Such a granularity of consent is extremely important in case if a processing activity involves multiple processing operations for more than one purpose. All processing activities carried out for the same purpose or purposes should be covered by consent, as outlined by Recital 32. When the processing has multiple purposes, consent should be given to all of them, respectively. Data subject must be granted the possibility to choose to accept or refuse the processing rather than having to consent to a "package deal". This is also closely related to the requirement of specificity.

A further element for a consent to be valid is that it must be "specific", that is, it can be given to "one or more specific purposes". Due to this condition, blanket consents and the use of catch-all phrases will result in invalidity.[130] In order to be specific, consent must be intelligible, that is, clear and precise about the scope and consequences of the processing.[131] Controllers must fulfil three conditions for providing a specific consent, namely the purpose specification, granularity, and clear separation of information related to obtaining consent from information about other matters.[132] To that end, Art. 7 (2) requires that in cases if the data subject's consent is given in the context of such a written declaration that also concerns different matters, the request for consent shall be presented in a manner which is "*clearly distinguishable from the other matters*". With respect to certain areas of scientific research, including certain analytics on Big Data, it should be noted that it is often not possible to fully set out the purpose of processing personal data at the time of the data collection. Therefore, the GDPR acknowledges 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*".[133] In research, consent for subsequent steps can be obtained before the next stage begins, provided that this is in line with the relevant ethical standards.[134] Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose.[135] The third element for a valid consent is that the consent must be "informed". The GDPR requires controllers to take significant steps in order to ensure that data subjects are provided with sufficient information. Data subjects must be provided with information prior to obtaining their consent. The two requirements data controllers must accomplish to ensure appropriate information are the quality of the information as well as the accessibility and visibility of the information. The first addresses the issue of what kind of information must be provided, while the second deals with the conditions on how to provide information. As far as the minimum content requirements are concerned, Recital 42 clarifies that "*for consent to be informed, the data subject should be aware at least of the identity of the controller and the purposes of the processing for which the personal data are intended*". In practice, the elements that are crucial for data subjects in order to understand what they give consent for are:

1. the controller's identity – in case of multiple/joint controllers, all controllers should be named
2. the purpose of each of the processing operations for which consent is sought
3. the type of data that will be collected and used
4. the existence of the right to withdraw consent
5. information about the use of the data for decisions based solely on automated processing, including profiling, in accordance with Art. 22 (2)

[130] Stefan Ernst, 'Die Einwilligung nach der Datenschutzgrundverordnung – Anmerkungen zur Definition nach Art. 4 Nr. 11 DS-GVO' [2017] ZD 110, 113
[131] see also Art. 7 GDPR; Gerald Spindler and Lukas Dalby, Art. 7 para. 10 in Spindler and Schuster (eds) (n 122)
[132] Art. 29 WP, WP 259 (n 123) 12
[133] R 33 GDPR
[134] Art. 29 WP, WP 259 (n 123) 28
[135] R 33 GDPR

6. the possible risks of data transfer to third countries in the absence of an adequacy decision and appropriate safeguards.[136]

Due to the sensitivity of the data the requirements for consent under Art. 9 (2) lit. a) are higher than those for a consent related to non-sensitive personal data.[137] The important additional requirement is the feature of expression. Namely, consent within the meaning of this provision must additionally be "***explicit***".[138] An implicit statement, implied declaration or an opt-out consent policy will not result in a valid consent.[139] Following recital 32, explicit consent does not mean that the consent cannot be given in electronic or oral form, which may, however, make it more difficult to provide evidence at a later stage.[140] For a consent to be explicit the consent form should reasonably include the designation of the exact category and type of data, beyond the general conditions of effective consent, in order to make the explicit risk of processing tangible for the data subject.[141]

However, healthcare providers acting as data controllers might not want to go down this road.[142] The reason for this is that a patient – doctor (hospital) relation is often a one-sided dependent one. In case dependency or power relationship is apparent in the relationship between the controller and the data subject in the run-up to the processing of health data, there is disagreement as to whether the criterion of voluntariness can nevertheless be met, or whether consent within the meaning of Art. 9 (2) lit. a) in conjunction with Art. 7 is ruled out in the absence of the fundamental necessity of equality.[143] Both R 43 and the Art. 29 WP refer to processing by public authority and processing in the employment context as clear imbalance of power.[144] Processing special categories of personal data by healthcare providers is not explicitly mentioned by the GDPR as an example for imbalance. On the other hand, voluntariness could be jeopardised if the people giving their consent are particularly dependent on the given service, especially in the health care sector.[145] However, the question of genuine freedom of choice requires a more detailled consideration. If it can be ensured with sufficient certainty in an individual case that the consent was granted voluntarily and in the interest of the data subject, there is rightly little reason why it should not be recognised as a legal basis for lawful processing. This can be supported by the wording of R 43, which negates freely given consent if in the light of all the circumstances of the particular case, it is unlikely that consent was given voluntarily.[146] When assessing the circumstances, account should be taken of the particularly high requirements, such as usefulness for the data subject, unequivocal reference to voluntariness.,[147] Accordingly, the patient has no real choice if he would have to fear

[136] Art. 29 WP, WP 259 (n 123) 13-14

[137] Council, 'Position (EU) No 6/2016 of the Council at first reading with the view to the adoption of a Regulation of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)' OJ C 159/86, available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:C:2016:159:FULL&from=GA accessed 12 December 2019

[138] Gerald Spindler, 'Big Data und Forschung mit Gesundheitsdaten in der gesetzlichen Krankenversicherung' (2016) Medizinrecht 34 9 691, 697; Härting, (n 78) [552]

[139] Alexander Schiff, Art. 9 para. 33 in Ehmann and Selmayr (eds) (n 40); Thilo Weichert, Art. 9 para. 47, in Kühling and Buchner (eds) (n 21)

[140] Gerald Spindler and Lukas Dalby, Art. 9 para. 7 in Spindler and Schuster (eds) (n 122)

[141] Kai-Uwe Plath, Art. 9 para. 13 in Plath (ed) (n 108)

[142] Anne Paschke, Datenschutz im Medizinsektor para. 58 in Louisa Specht and Reto Mantz (eds) *Handbuch Europäisches und deutsches Datenschutzrecht* (C.H.Beck 2019)

[143] Eike Michael Frenzel, Art. 9 para. 24, in Paal and Pauly (eds) (n 104)

[144] R 43 s 1 GDPR; Art. 29 WP, WP 259 (n 123) 6-7; with emphasis on processing by public authority based on fear of future remedies Eike Michael Frenzel, Art. 9 para. 24, in Paal and Pauly (eds) (n 104)

[145] Albert Ingold, Art. 7 para. 27 in Sydow (ed) (n 108); Dirk Heckmann and Anne Paschke, Art. 7 para. 52 in Ehmann and Selmayr (eds) (n 40)

[146] Gerald Spindler and Lukas Dalby, Art. 9 para. 8 in Spindler and Schuster (eds) (n 122); R 43 s 1 GDPR

[147] Thilo Weichert, Art. 9 para. 51, in Jürgen Kühling Benedikt Buchner (eds) (n 21)

disadvantages if his consent were refused.[148] The same is true if the treatment is made dependent on the consent to process data not relevant to the treatment.[149] Generally it can be said that consent should not be considered voluntary if compulsion, pressure or inability prevents "*to exercise free will*".[150] On the other hand, healthcare providers are not per se banned from using consent as lawful basis, but they need to take measures to ensure that the patient does not feel pressure to give consent and to alleviate concerns about the consequences of refusing to give it.[151]

2.2.2.2.2. Processing necessary for the purposes of preventative or occupational medicine, for assessing the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or management of health or social care systems and services on the basis of Union or Member State law or a contract with a health professional– Art. 9 (2) lit h)

As discussed above, in view of the fact that often the absence of a power relation is difficult to demonstrate, data controllers might want to consider a different lawful basis. Art. 9 (2) lit h) might be an alternative. Together Art. 9 (2) lit h) in accordance with Recital 53 enables the processing of sensitive data for medical diagnoses and treatment and the management of health or social care systems and services, if necessary.[152] This provision ensures that data protection does not prevent effective medical treatment.[153]

R 53 clearly states, that this exemption is applicable "*only where necessary to achieve those purposes for the benefit of natural persons and society as a whole*".[154] As the recital indicates the processing can be based on this provision in case of a favourable outcome of a necessity test. The processing must therefore serve both the protection of the data subject and the general interest of providing adequate healthcare.[155] This provision allows exceptions to the general prohibition of processing initially for the purpose of preventive medicine. According to R 53, health care within the frames of this provision should also serve to ensure and monitor health and health treats.[156] Processing within the provision of "*medical diagnoses and treatment*" must relate to the treatment itself, or to the management and administrative issues of that.[157] Medical care includes treatment and diagnostics in a broader sense and related infrastructural measures such as bookkeeping, as well as health care including the development and implementation of care concepts.[158]

Paragraph 2 lit h) first of all requires a legal basis for the processing in Union or Member State[159] law or in the form of a contract with a healthcare professional. This requirement implies processing "on the basis of a contract", which presupposes the existence of a contract already in place.[160] The medical services provided are usually based on a contract between the individual concerned and the medical

[148] Anne Paschke, Datenschutz im Medizinsektor para. 58 in Louisa Specht and Reto Mantz (eds) (n 142)

[149] ibid, para. 58

[150] Art. 29 WP, WP 259 (n 123)

[151] ICO, 'When is consent appropriate?' available at https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/consent/when-is-consent-appropriate/ accessed 23 December 2019

[152] Härting, (n 78) [543]

[153] David Kampert, Art. 9 para. 43 in Sydow (ed) (n 108)

[154] R 53 s 1

[155] Britta Alexandra Mester, Art. 9 para. 31 in Taeger and Gabel (eds) (n 23)

[156] R 53 s 1

[157] Sebastian Schulz, Art. 9 para. 35 in Gola (ed) (n 39)

[158] Gerald Spindler and Lukas Dalby, Art. 9 para. 18 in Spindler and Schuster (eds (n 122); Thilo Weichert, Art. 9 para. 104 in Kühling and Buchner (eds) (n 21)

[159] e. g. § 22 section 1 Nr. 1 lit. b) of the Federal Data Protection Act of Germany (FDPA)

[160] Thomas Petri, Art. 9 para. 85 in Simitis, Hornung and Spiecker (eds) (n 9)

service provider. These contracts between patients and healthcare professionals are called treatment contracts.[161] The general contractual regulations apply to this contract, including on general terms and conditions or on other protection of the weaker party of the contract.[162] Given that the higher protection of health data serves the same purpose, namely the protection of the party in the weaker position, it cannot be said that contractual regulations impede the right to informational self-determination and patient privacy.[163]

The other scenario regulated in Art. 9 (2) lit h) is processing for management of health and social care systems, which includes the entire organisational institution for performing of health services. In contrast to the Directive, the new concept of the GDPR is broader and encompasses procedures and contracts under private law as well, such as coverage of costs by insurance companies.[164]

An important requirement for this is, according to Art. 9 (3), that the data "*are processed by or under the responsibility of a professional subject to the obligation of professional secrecy under Union or Member State law or rules established by national competent bodies or by another person also subject to an obligation of secrecy under Union or Member State law or rules established by national competent bodies*". This provision narrows down the scope of Art. 9 (2) lit. h) as it stipulates that the processing of sensitive data for the purposes indicated here is only permitted under the responsibility of qualified personnel who are subject to professional secrecy or a comparable obligation of secrecy.[165]

2.2.2.2.3. Processing necessary for archiving purposes in the public interest, or scientific and historical research purposes or statistical purposes in accordance with Article 89 (1)– Art. 9 (2) lit j)

With establishing the exception of Art. 9 (2) lit j) the GDPR introduced a new provision that has notable positive implications especially in the medical research sector, since it provides additional legitimate ground on which sensitive personal data may lawfully be processed. It should be pointed out that Art. 9 (2) lit j) only refers to independent research. An external influence on the scientific findings or deferring those findings to sheer economic purposes must be precluded. On the other hand, the fact that a project is financed by third-party-funds does not necessarily exclude the applicability of Art. 9 (2) lit j).[166] On the other hand, it covers both research proposals initiated and carried out by the same controller and external projects, i. e. where the access to sensitive data is granted by the controller, but the implementation of the project, the analysis of those data is outsourced to one or more processors.[167]

2.2.3. Determination of purpose – secondary research purposes

The principle of purpose limitation is one of the founding principles of the current data protection framework. Purpose limitation is a prerequisite for predictable data management as it means basically the prohibition of aimless data collection. The adherence to this principle shall be monitored and validated at all stages of the processing of personal data.

According to Art. 5 (1) lit b), personal data "*shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or*

[161] Sebastian Schulz, Art. 9 para. 38 in Gola (ed) (n 39)

[162] e. g. § 305 ff and 630a ff of the German Civil Code, respectively

[163] Thilo Weichert, Art. 9 para. 103 in Kühling and Buchner (eds) (n 21)

[164] ibid, para. 105 ff

[165] Kai-Uwe Plath, Art. 9 para. 24 in Plath (ed) (n 108); R 53 s 1; Silke Jandt, 'Smart Health. Wird die DSGVO den dynamischen Herausforderungen gerecht?' [2016] DuD 571, 574

[166] ibid, para. 129

[167] Sebastian Schulz, Art. 9 para. 43 in Gola (ed) (n 39)

statistical purposes shall, in accordance be explicit and legitimate and determined at the time of the collection of the personal data".

The prerequisites 'specified' and 'explicit' not only serve the purpose limitation principle, but are also closely connected with transparency, data minimisation regulated in Art. 5 (1) lit. a) and c) respectively and with protection of the data subject's rights. For the sake of user control, the purpose should be precise and clear enough to predict how and to what extent the controller handles the data in question. For a purpose to be specified, it must be sufficiently defined to delimit the scope of the processing operation. Secondly, the purpose must be unambiguous and clearly expressed, without any hidden agenda, such as secret algorithms or hidden profiling.[168] Purpose specification requires an internal assessment carried out by the data controller, prior to, and in an event, not later than, the time when the collection of personal data occurs.[169] Expressing the purpose in writing and adequate documentation will also help the controller to demonstrate the compliance with the requirements of Art. 5 (1) b). In addition, the purpose of the processing must be legitimate.

The second building block of this principle is the compatible use of personal data in case of further processing.[170] The legislator opted for a double negation on this matter by stating that personal data "*shall not be further processed in a manner that is incompatible with those purposes*".[171] It is important to notice that a different purpose does not necessarily and automatically results in incompatibility, this needs to be assessed on a case-by-case basis.

If further processing is permitted, no legal basis other than the legal basis of the original data collection is required,[172] since the legal basis of the Union or national state can also be used for further processing.[173] However, this does not release the data controller from his obligation to comply with all the requirements of the lawfulness of the original processing, such as the fulfilment of at least one of the "conditions of lawfulness" pursuant to Art. 6 para. 1 lit. a-e.[174] Compliance with the data protection principles pursuant to Art. 5 and the existence of an appropriate legitimate basis pursuant to Art. 6 (1) shall therefore be regarded separately.[175]

According to Art. 6 para. 4 in conjunction with Art. 5 para. 1 lit. b) a compatibility assessment of the purposes, the so-called compatibility test, can be waived in three cases. First, further processing for other newly introduced purposes is not permitted if the original data collection was based exclusively on the consent of the data subject. The reason for this is the irrevocable conditions of permissibility of consent, namely the fact that the purposes pursued must always be known and specified in the course of an informed consent, so that again there is no scope for "purposes other than the original" that have not been agreed upon. Following this logic, any change of purpose must be legitimised by a new consent directed towards this.[176] Secondly, no compatibility assessment is necessary if the processing is based on Union or national law and has a socially necessary and proportionate protective function vis-à-vis important data, in particular those laid down in Art. 23 (1).[177] Last but not least, Art. 5 (1) lit. b) states

[168] Art. 29 WP, WP 203 (n 46) 69

[169] Tobias Herbst, Art. 5 para. 31 in Kühling and Buchner (eds) (n 21)

[170] R 50 s 1

[171] It is unclear, what exactly qualifies as "further processing". The Art. 29 WP takes the view that any processing following collection must be considered "further processing" and must, therefore meet the requirement of compatibility. Art. 29 WP, WP 203 (n 46) 21

[172] R 50 s 2

[173] Jürgen Kühling and Mario Martini, 'Die Datenschutz-Grundverordnung: Revolution oder Evolution im europäischen und deutschen Datenschutzrecht?' [2016] Europäische Zeitschrift für Wirtschaftsrecht 448, 451; Manfred Monreal, 'Weiterverarbeitung nach einer Zweckänderung in der DS-GVO' [2016] ZD 507, 512

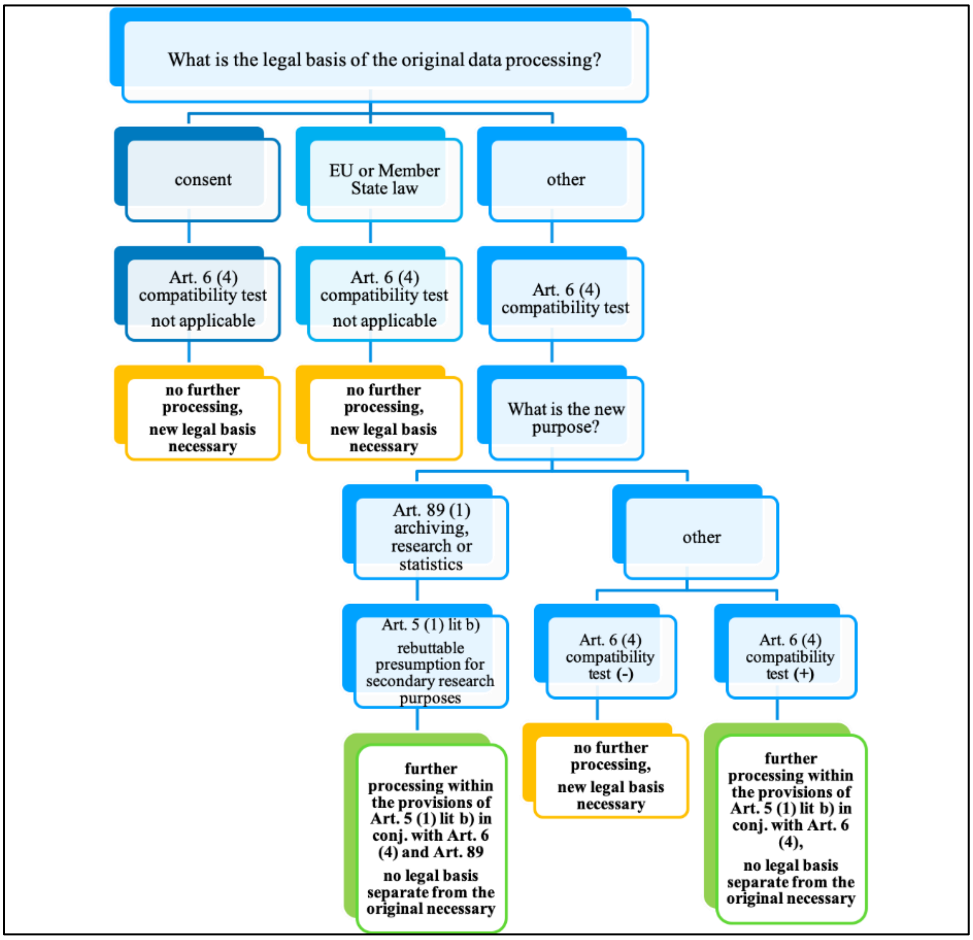
[174] Art. 29 WP, WP 203 (n 46) 37

[175]Benedikt Buchner and Thomas Petri, Art. 6 para. 184 in Kühling and Buchner (eds) (n 21)

[176] Monreal (n 173) 510

[177] Art. 23 (1) mentions, inter alia, the guarantee of public security (lit. c), national defence (lit. b) or the protection of other overriding public interest objectives in the monetary, budgetary and fiscal field or in the context of public health and social security (lit. e).

that further processing for archival purposes of public interest, for scientific or historical research purposes and for statistical purposes in accordance with Art. 89 (1) – taking into account appropriate guarantees such as TOMs – shall not be regarded as incompatible with the original purposes and they shall be considered compatible and lawful further purposes.[178] Whether this legal presumption of '*non-incompatibility*' is to be equated with a determined compatibility and whether this releases the data controller from carrying out a compatibility test is not clear. Although the benefit of a test "with a foreseeable outcome" prima facie seems rather low, further processing for the benefit of privileged purposes also requires a compatibility assessment. Art. 5 (1) lit. b) is designed as a legal presumption which indeed prejudices the result of the compatibility test, but this legislative evaluation can only flow into a compatibility assessment, if it is allowed to be carried out at all.[179]


What is the legal basis of the original data processing?
consent
EU or Member State law
other
Art. 6 (4) compatibility test not applicable
Art. 6 (4) compatibility test not applicable
Art. 6 (4) compatibility test
no further processing, new legal basis necessary
no further processing, new legal basis necessary
What is the new purpose?
Art. 89 (1) archiving, research or statistics
other
Art. 5 (1) lit b) rebuttable presumption for secondary research purposes
Art. 6 (4) compatibility test (-)
Art. 6 (4) compatibility test (+)
further processing within the provisions of Art. 5 (1) lit b) in conj. with Art. 6 (4) and Art. 89
no legal basis separate from the original necessary
no further processing, new legal basis necessary
further processing within the provisions of Art. 5 (1) lit b) in conj. with Art. 6 (4),
no legal basis separate from the original necessary

3. Figure Repurposing the original data processing pursuant to Art. 5 (1) lit b) in conj. with Art. 6 (4) and Art. 89 (1) GDPR

As to special categories of personal data, the relationship between Art. 9 and Art. 6 (4) - change of purpose – has been controversial in the scientific discourse. However, as Art. 6 (4) lit. c) refers to Art. 9, it can be assumed that Art. 9 has no restrictive effect with regard to that paragraph. A change of purpose is therefore also conceivable for the processing of special categories of personal data but is subject to a special justification.

[178] in accordance with R 50 s 4

[179] Gerald Spindler and Lukas Dalby, Art. 6 para. 24 in Spindler and Schuster (eds) (n 122)

2.2.4. Division of roles and outsourcing

This is a very much relevant problem related to the cases discussed below, since the roles different parties play during the processing come with different responsibilities, obligations and liability for each party.[180] Since SODA offers a cloud-ready data analytics tool toll, it is pivotal to properly classify the parties of the outsourcing scheme within the data protection terminology. There are multiple stakeholders taking part in the processing activities along the entire Big Data value chain, not every motivated party act as data controller or data processor. Concerning outsourcing and cloud computing, where typically multiple actors are involved, and it is not always unambiguous who acts as a data controller, data processor, recipient or third party in terms of the GDPR.

One way to investigate this issue is to clarify what data controller and data processor mean according to the GDPR and try to match these abstract definitions with the terms commonly used to describe the players in a cloud computing setting.[181]

2.2.4.1. Data Controller and Data Processor

In every situation when someone processes personal data, it does so either as data controller or data processor.

2.2.4.1.1. Data Controller

The first option is that an organisation has its own resources to carry out the data processing. A scenario, where a party processes the data itself without outsourcing it to data processors is called in-house processing, provided that they match the criteria of the definition for data controller. Art. 4 Nr. 7 GDPR defines data controller as:

> "*the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data*."

The rule of a controller is not assigned exclusively to one party, rather it shifts from one party to another depending on the specifics of the processing, and this allocation of role need to be interpreted broadly, in accordance to the specific context of processing.[182] This pragmatic approach of data controller fits to the dynamic changes and diversity of data processing settings.[183]

The two building blocks of the definition are 1) defining the purposes and means and 2) alone or jointly with others. Pursuant to Art. 24 (1) the data controller must continuously evaluate the nature, scope, circumstances and purpose of the data processing, including the risks to the rights and freedoms of the data subject, and take appropriate technical and organisational measures to ensure (and subsequently provide evidence of) that the processing is carried out in accordance with the regulations.[184]

[180] cf. Härting (n 78) [570 ff]

[181] Note that due to didactic reasons this section examines data controllers and data processors on a conceptual level and focuses on the definitions and distinctions between them. The different obligations and liabilities are not presented here, for a detailed discussion about those see Gerald Spindler and Anna Zsófia Horváth, 'D3.1 General Legal Aspects' (n 86)

[182] ECJ decision of 13/05/2014 – C-131/12 *Google Spain SL, Google Inc. v Agencia Española de Protección ded Datos (AEDP), Mario Costeja González* ECLI:EU:C:2014:317 para. 34

[183] Art. 29 WP Opinion 01/2010 on the concepts of "controller" and "processor" [2010] WP 169, 9 available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2010/wp169_en.pdf accessed 21 November 2019

[184] Gerald Spindler and Lukas Dalby, Art. 4 Nr. 7 para. 18 in Spindler and Schuster (eds) (n 122)

Accordingly, the data controller has to be the 'lord of the data' who decides the objectives of the processing and determines the means by which these objectives are to be achieved.[185] The term "*means*" is interpreted broadly in this context and encompasses a wide range of methods. Albeit the controller must remain the decisive entity, in many cases it would not be feasible to put the burden of selecting the most appropriate way of processing entirely on them. To solve this problem, data protection rules grant data controllers delegable competences when it comes to the technical and organisational measures of the processing. This shall not apply where essential elements of the means are concerned, such as questions regarding the type of the data, the duration of the processing or access control.[186] It is specifically for situations where controllers may not be in the position to determine the exact means on their own, this obligation can be delegated to some extent.[187] As a result, processors may have some degree of discretion. Bearing in mind this right to delegate it shall be noted that a mere power of decision in respect of the means of processing is normally not sufficient to classify an actor as data controller.[188] On the other hand, if the processor takes over the decision regarding the details on how the data analysis should be carried out, the chosen means should represent a "reasonable way" of achieving the purposes set out by the controller.[189]

The second element of the definition is that controllers may act "*alone or jointly with others*". The collaboration between actors involved in processing of personal data is not necessarily a controller – processor relation, it is also possible that multiple actors interact in the processing of personal data as controllers.[190]A complete assessment of all specific circumstances is advised for each actor in order to decide whether decisions should be made jointly as a separate controller.[191]

2.2.4.1.2. Joint Controllers

The definition in Art. 4 opens the door for joint controllership arrangements.[192] Joint controllership is regulated in Art. 26, where the GDPR defines this formation as:

"*two or more controllers jointly determine the purposes and means of processing, they shall be joint controllers. [...]*"

This definition does not contain any new substantial requirements in terms of content compared with the legal definition of the data controller in Art. 4 Nr. 7.[193] Art. 26 (2) defines joint controllership through procedural obligations, namely, that they oblige to determine the distribution of control in a transparent manner by means of an arrangement between them and provide a summary of the arrangement available to the data subject. This clarification is of utmost importance firstly because of the clear allocation of responsibilities in order to comply, secondly concerning liability issues.[194]

[185] Härting (n 78) [571]
[186] Art. 29 WP, WP 169 (n 183) 17
[187] Note however, that the competences related to the decision about the purpose of the processing as well as to the substantial elements of the means are non-delegable.
[188] Bernd Wagner, 'Disruption der Verantwortlichkeit, private Nutzer als datenschutzrechtliche Verantwortliche im Internet of things' [2018] ZD 307, 309
[189] Art. 29 WP, WP 169 (n 183) 14
[190] Brendan Van Alsenoy, 'Liability under EU Data Protection Law: From Directive 95/46 to the General Data Protection Regulation' (2016) JIPITEC 7 271, 279 available at https://www.jipitec.eu/issues/jipitec-7-3-2016/4506 accessed 28 November 2019
[191] Art. 29 WP, WP 169 (n 183) 18
[192] Jan Philipp Albrecht and Florian Jotzo 'Allgemeine Bestimmungen, Begriffsbestimmungen' section 3 A para. 7 in Jan Philipp Albrecht and Florian Jotzo, *Das neue Datenschutzrecht der EU* (Nomos, 2017)
[193] Thomas Petri, Art. 26 para. 5 in Simitis, Hornung and Spiecker (eds) (n 9)
[194] Art. 29 WP WP 169 (n 183) 22

The ECJ has addressed the question of joint controllers recently in its '*Fan page*' and '*Jehovah's witnesses*' decisions.[195] The Court ruled in the Fan page decision that an administrator of a Facebook fan page is jointly responsible with Facebook for processing the personal data of the visitors to its page.[196] According to the Court's arguments it is not decisive for joint control that the fan page administrator has no influence on and/or access to data processing by the social media site.[197] Instead, the Court introduced a term "*parameters*" in its reasoning, arguing that if an actor has an opportunity to influence the processing through parameters in particular "*on […] the objectives of managing and promoting its activities*" this actor meets the criteria of the joint controller at least regarding to this particular processing activity.[198] This conclusion leads to a very watered down notion of joint controllership.[199]

Remarkable is a clarification by the ECJ on the weighting of responsibility in the case of joint controllership. The ECJ's argumentation on joint responsibility in the Fan page decision is followed by the interpretation that the existence of joint responsibility does not mean that the different actors involved in the processing of personal data automatically share equal responsibility. The Court is of the opinion that joint controllership does not necessarily mean an equivalent responsibility.[200] The ECJ emphasises once again in the Jehovah's Witness judgment that it is possible for actors – data controllers – to be involved in the processing of personal data at different stages and to different extents, so that the degree of responsibility of each of them must be assessed taking into account all relevant circumstances of the individual case.[201]

The most recent landmark case of the CJEU dealing with joint controllership is another step towards a broad interpretation of this formation. In this so called 'Fashion ID' judgement the Court held that an operator of a website embedding third-party features, specifically in this case a Facebook 'like'-button, is jointly responsible with Facebook within the context of this activity.[202] The Court refers to the Fan page and Jehovah's Witnesses judgements as it argues that under the broad notion of data controller actors who have influence over the processing of the personal data and participate in the determination of means and purposes could be regarded as joint controllers.[203] The finding that joint controllership is possible despite of the fact that not all parties have access to the same dataset is also consistent with prior statements as well as with the position of the Art. 29 WP.[204] A new element of the ECJ's reasoning is however that joint controllership can exist for specific phases of the data processing and that a given actor cannot be considered to be a controller "*in the context of operations that precede or are subsequent in the overall chain of processing for which […] does not determine either purposes or means*".[205] In this particular case the Court concluded that the operator of a website embedding the like button and Facebook are joint controllers in respect of the collection and transmission of personal data, but not in respect of subsequent processing by the social media provider. The take-away of this conclusion is the reasoning that where the processing of personal data consists of a number of subsequent operations,

[195] ECJ decision of 5/06/2018 – C-210/16 *Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig-Holstein GmbH* ECLI:EU:C:2018:388; ECJ decision of 10/07/2018 – C-25/17 *Korkein hallinto-oikeus v Tietosuojavaltuutettu* ECLI:EU:C:2018:551; Flemming Moos, 'Update Datenschutz' [2018] DSRITB 259, 259ff

[196] ECJ C-210/16 (n 195) [75]

[197] Ibid [38]; Niko Härting and Patrick Gössling, 'Gemeinsame Verantwortlichkeit bei einer Facebook-Fanpage' [2018] NJW 2523, 2424

[198] ECJ C-210/16 (n 195) [36, 69]

[199] Judith Nink, Art. 26 para. 7 in Spindler and Schuster (eds) (n 122)

[200] ECJ C-210/16 (n 195) [43]

[201] ECJ C-25/17 (n 195) [66]

[202] ECJ decision of 29/07/2019 – C-40/17 *Fashion ID GmbH & Co. KG v Verbraucherzentrale NRW eV.* ECLI:EU:C:2019:629

[203] Ibid [66-68]

[204] Ibid [69]; ECJ C-210/16 (n 195) [38]; ECJ 25/17 (n 195) [69]; Art. 29 WP, WP 169 (n 183) 22 „*having access to data is not an essential condition to be a controller*"

[205] ECJ C-40/17 (n 202) [74]

joint controllership can be established to one or more of these stages respectively, depending on the influence and joint decision-making power the parties exert over a given operation.[206]

2.2.4.1.3. Data Processor

Most of the time data controllers lack the resources and/or know-how to carry out the data processing themselves. Therefore, they need to rely on outside data processors and engage in a controller-processor relation according to Art. 28.
According to Art. 4 Nr. 8 data processor is

> "*a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller*."

The two key conditions to qualify as a processor are firstly to be a separate legal entity with respect to the controller[207], secondly, processing personal data on the controller's behalf.[208] The lawfulness of his processing activity depends entirely on the mandate given by the controller.[209] This has a dual meaning. Firstly, processors must act in the controller's best interest, and not for their own purposes.[210] Secondly, they are bound to the controller's instructions, that is to say, "*shall not process data except on instructions from the controller*".[211]

2.2.4.2. Outsourcing of the processing

As mentioned above, data controllers often do not dispose of the necessary means and resources to process large-scale of data on their own. The controller does not necessarily have to be the entity actually processing the data. On the contrary, companies whose main business is outside the IT-sector tend to outsource data processing.
When it comes to outsourcing, a distinction is often made between "classic" outsourcing and cloud computing. Cloud computing generally includes services that are based on virtualization, resource pooling and multi-tenancy systems, whereas "classic" outsourcing is characterized by the provision of dedicated resources. However, the distinction is not always clear because, on the one hand, dedicated resources are also used in the private cloud and, on the other hand, providers are increasingly working with technologies such as virtualization even in the supposedly "classic" outsourcing environment.[212]
In any case, outsourcing data processing activities results in the emergence of data processing chains. This is also true for cloud computing, since what the cloud offers have in common is that there is regularly not one entity on the contractor – data processor - side. There are two main models of processing chains in case of outsourcing. The chain always starts with the data controller. It is '*on the top*'. Depending on the number of and agreement with the processors the chain can have either a linear vertical or a sort of 'pyramid' shape.
The first variation appears where the first processor in the chain has a direct contractual relation with the data controller and engages further sub-processors to carry out specific aspects of the processing.

[206] cf. ibid [69 ff]
[207] otherwise it would be in-house processing
[208] Since the processor handles the data exclusively on behalf of the controller based on his instructions, sometimes this constellation is referred to as 'order-processing'
[209] Art. 29 WP, WP 169 (n 183) 22
[210] Should the processor either reach further in processing the data as to what the controller intended to or use the data for its own purposes, it will be considered as controller relating to that processing operations, Art. 28 (10) GDPR
[211] Art. 29 GDPR
[212] Thomas Thalhofer and Konrad Żdanowiecki, 'Outsourcing-Verträge' para. 1 in Astrid Auer-Reinsdorff and Isabell Conrad (eds) *Handbuch IT- und Datenschutzrecht* (3th edn C.H.Beck 2019)

Data processors are generally allowed to do so, provided that this occurs with the knowledge and authorisation of the data controller.[213] Pursuant to Art. 28 (2) GDPR the processor „*shall not engage another processor without prior specific or general written authorisation of the controller. In the case of general written authorisation, the processor shall inform the controller of any intended changes concerning the addition or replacement of other processors, thereby giving the controller the opportunity to object to such changes"*. This should ensure, that the controller stays the determining party. Article 28 (4) elaborates on the case where a processor engages a sub-processor stating that "*the same data protection obligations as set out in the contract or other legal act between the controller and the processor […] shall be imposed on that other processor by way of a contract or other legal act under Union or Member State law, in particular providing sufficient guarantees to implement appropriate technical and organisational measures […]*. In this case the instructions of the controller run down the chain to processor and sub-processor top-down.

It is also common, that different phases of the processing are outsourced to different processors. It is conceivable for two separate entities to be data processors of the same data, e. g. one of them runs the analytics whereas the other stores the data – both are data processors in relation to the same dataset. The aforementioned second variation occurs when all these data processors have a direct contractual relation with the controller. Furthermore, a hybrid model by the combination of the two models is possible as well.

Generally speaking, outsourcing by cloud computing is a complex process consisting of uploading, transferring, combining and storing of data. As a result of the basic functioning of cloud computing, there is some form of data processing happening in the cloud constantly. The processing is either 'active' where something happens to the data in motion or 'passive' where the data is not used actively but is stored as data at rest.[214]

In cloud computing, the cloud client uses a structure of the service provider where the availability of the data is of primary importance for the client. If a cloud service provider processes personal data, the question arises as to whether this can be referred to as order data processing.[215] Apart from the agreement between cloud client and cloud provider the answer to this question depends on two additional factors:

1. cloud service provision model (IaaS, PaaS, SaaS)
2. design of the cloud (public cloud, private cloud, hybrid cloud).[216]

It is not without difficulty to bring into accordance the abstract defined terms of the GDPR with the terms used for describing parties in cloud computing. The GDPR never explicitly states that a cloud service provider is a data processor by definition. Notwithstanding it appears so that the Art. 29 WP treats cloud providers as data processors rather than controllers or as a third party.[217] This argument does not get any less valid because of the mere fact that the cloud service provider usually selects the means to be used for processing on the basis of his technical expertise. This does not change his classification as a contract processor, since the party acting as data controller may transfer this choice of means to the order data processor in the order agreement.[218]

[213] Isabell Conrad and others, 'Cloud Computing Verträge' para. 201 in Auer-Reinsdorff and Conrad (eds) (n 212)

[214] Nate Lord, 'Data Protection: Data in Transit vs. Data At Rest' (*Data Insider – The Guardian's digital blog*, 15 July 2019) available at https://digitalguardian.com/blog/data-protection-data-in-transit-vs-data-at-rest accessed 28 November 2019

[215] Rudi Kramer Art. 18 para. 16 in Gierschmann and others (eds) (n 81)

[216] Isabell Conrad and others, 'Cloud Computing Verträge' para. 208 in Auer-Reinsdorff and Conrad (eds) (n 212)

[217] Art. 29 WP Opinion 05/2012 on cloud computing [2012] WP 196 available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2012/wp196_en.pdf accessed 29 November 2019

[218] see 2.2.4.1.

Whit that said it seems that the prevailing majority considers cloud service providers to be data processors.[219] This would lead to the fact that they were subject to all requirements and obligations regarding data processors of the GDPR.[220] Albeit this view solves the issue at first glance, such blanket classification ignores the argument that a differentiated consideration should be made as to whether the cloud service provider merely provides the infrastructure (IaaS) or whether additional services are provided. In the case of the mere provision of an infrastructure, typically no user-related personal data would be processed by the cloud provider. Therefore, at least in case the IaaS provider does not have access to the cloud client's personal data, the conditions of data processing pursuant to Art. 4 Nr 8 and Art. 28 GDPR would not be duly fulfilled.[221]

The answer to this question has a particularly significant influence on the evaluation of outsourcing of the processing of anonymised data. De-identifying personal data in a way that it could be considered anonymised data with regard to the conditions set by the absolute approach with relative elements could arguably take cloud providers out from under the scope of data protection legislation.[222] For IaaS providers this could be true regardless and despite of the fact that the data should be considered personal for SaaS providers using IaaS infrastructure, assuming that the IaaS provider cannot re-identify the encrypted or otherwise anonymised dataset.[223]

[219] Art. 29 WP, WP 196 (n 217) 8-9; Hans Markus Wulf and Clemens Burgenmeister, 'Industrie 4.0 in der Logistik – Rechtliche Hürden beim Einsatz neuer Vernetzungs-Technologien (2015) CR 404, 410-411; Mark Webber, 'The GDPR's impact on the cloud service provider as a processor' [2017] Privacy & Data Protection 16 4, available at https://www.fieldfisher.com/media/3993765/the-gdprs-impact-on-the-cloud-service-provider-as-a-processor-mark-webber-privacy-data-protection.pdf accessed 11 December 2019

[220] Kuan Hon, Christopher Millard and Ian Walden, 'Who is responsible for 'personal data' in cloud computing? – The cloud of unknowing, Part 2' (2012) IDPL Vol 2 No. 1 7

[221] Isabell Conrad and others, 'Cloud Computing Verträge' para. 208 in Auer-Reinsdorff and Conrad (eds) (n 212)

[222] Kuan Hon, Christopher Millard and Ian Walden, 'Who is responsible for 'personal data' in cloud computing? – The cloud of unknowing' (2011) IDPL Vol 1 No. 4 211, 219

[223] ibid, 219

3. Use Cases

This section examines three use cases in the medical domain and discusses how secure multi-party computation enables data analytics on data from multiple providers.[224] The technical solution presented in each case resembles anonymisation using a trusted third party. However, and this is one of the most important innovations of MPC, the role of the trusted third party, who traditionally has an arrangement with the involved organisations, is taken over by the software itself. As a result, a 'physical' third party becomes dispensable. This setting is perfectly suitable for collaborative projects.

In each case, as a general rule, datasets contain different kind of personal data of different individuals. When exchanging their data, the actors must guarantee that the individuals will not be identified from the combined datasets. They have several ways to achieve successful analysis without compromising security. Data must either undergo anonymisation where, as a result the identifiable attribute is completely stripped from the natural person. In this case data cease to be personal data and may be processed as other feature data. The other method is to treat the data with secure cryptographic techniques which are compliant with the GDPR. A new dataset can be generated as an output, which contains the results of the analysis without allowing for personal identification. In the end, according to the specific research agreement or other instrument the anonymous output data can be used for further analysis or published in several ways.[225]

3.1. Case 1: Intra-sector internal MPC

The first case presents an intra-sector, internal MPC analysis where three hospitals share and analyse patient data to compare, predict and improve treatment outcomes with own resources. The hospitals taking part in the research share data of patients they have treated with diabetes. The research focuses on prediction of population health, for instance to predict the risk of other comorbidities within a given timespan for patients.

The hospitals perform joint analysis using data from different data providers in the same sector. Each of the participating hospitals is in fund of the necessary resources, including private cloud platforms. They run the MPC software using their own infrastructure. Using MPC, the model can be built without hospitals needing to share any data about their patients.

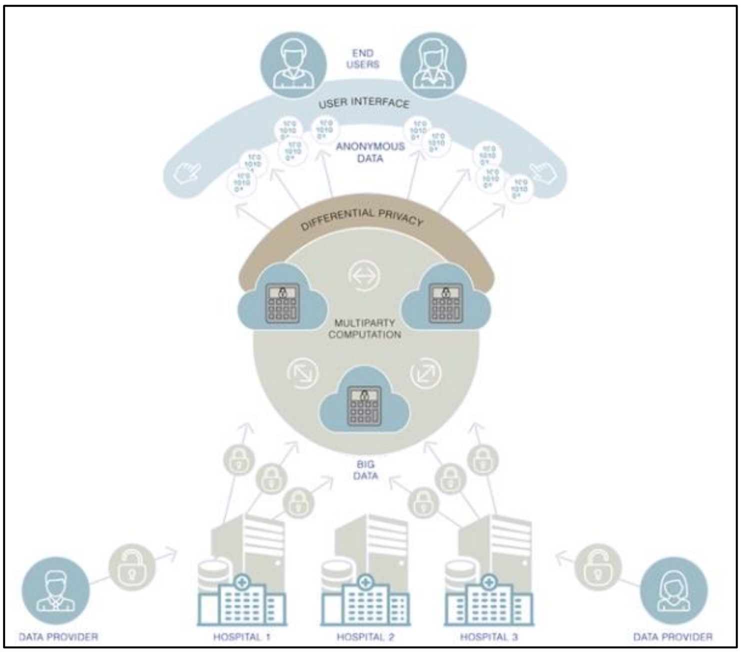
3.1.1. Functioning

SODA offers a complex data analytics system using secret sharing, multi-party computation and differential privacy. This is a cloud-ready data analysis system for secure processing of confidential information. By its design, it provides security without the risk of an insider attack. The raw data never leaves the hand of the "owner" on the input level, only the final results of the computation are shared with the partners. It provides privacy because private information can be processed without compromising the data subject's rights and convenience of use.

In order to guarantee efficient functioning and uncompromised security, three servers must be deployed by organisations that do not collude.

[224] With minor modifications the cases discussed here correspond to the cases presented as SODA pilot application cases in the conference paper: Meilof Veeningen and others, 'Enabling Analytics on Sensitive Medical Data with Secure Multi-Party Computation' [29th Medical Informatics Europe Conference Gothenburg Sweden 24-26 April 2018] available at http://ebooks.iospress.nl/volumearticle/48757 accessed 26 November 2019

[225] cf. ICO, 'Anonymisation: managing data protection risk – code of practice' (2012) ch 8 42 available at https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf accessed 22 November 2019



4. Figure SODA data processing model

3.1.1.1. Secret Sharing

Before the data is stored on different servers, it has to be divided, so that the random shares can be sent to the computing servers. This process is done over personal data in plaintext. Secret sharing is used at the pre-processing phase and prepares the data before the MPC analysis. In this application patient data can be split into pieces and these data shards can be stored in different databases on different servers for privacy-preserving purposes. In this particular case each party will receive one share of every secret value. The original secret can only be reconstructed by collecting all the shares of a value.[226]

3.1.1.2. Multi-Party Computation

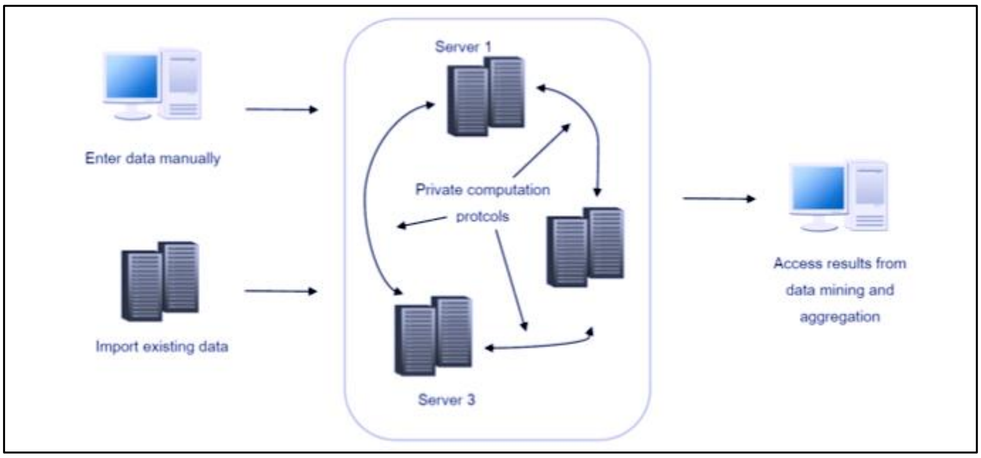
Secure multi-party computation refers to a field of cryptography that deals with protocols involving two or more participants who want to mutually compute a useful result.[227] Every party provides an input value and learns only the result of their own individual value so that nobody is able to access all the information.[228] In the case of data aggregation algorithms, it is generally not possible to learn the inputs of other parties from the result.[229] Figure 4 shows the data storage process with three servers. The data is collected from the users or exists already on other servers and is sent to the three servers. A data donor distributes the data into shards using secret-sharing and sends one random share of each value to a single server.

[226] Dan Bogdanov, Sharemind: programmable secure computations with practical applications (University of Tartu Press 2013) 34

[227] Claudio Orlandi, 'Is Multiparty Computation Any Good in Practice?' [Conference: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2011 Prague Czech Republic 22-27 May, 2011] available at https://www.cs.au.dk/~orlandi/icassp-draft.pdf accessed 29 November 2019

[228] Liina Kamm and Jan Willemson, 'Secure Floating-Point Arithmetic and Private Satellite Collision Analysis' (2015) 14/6 International Journal of Information Security 2

[229] ibid, 2



5. Figure General MPC functioning with three servers deployed by three institutions the information is divided between the three and every one of them receives a part of the information and works with it. At the end every part sends his results back to the client[230]

Since this time MPC is used to compare information from three entities in such a way that no one knows the other's values, both entities function as data donors. It could be necessary that one donor specifies what kind of information the other donor has to provide from its database; for example, the medical IDs of the persons whose data is about to be compared.

The separation of servers between input donors and processing servers is useful as it does not force every party to run secure multiparty computation protocols.[231]

After the data has been transmitted and stored, the server can perform computations on the shared data; however, the server does not share the information with other servers. This is done so that none of them can reconstruct the input values.[232] The number of three servers is used for efficiency and is needed to guarantee the security during the computation; otherwise, it would be too easy to reconstruct the data. Moreover, an increase of servers reduces any risk of collusions. Secure multiparty computation protocols specify which messages the server should exchange in order to compute new shares of the value that correspond to the result.[233] After finishing the computation, the results of the servers are transmitted and published to the client of the computation (the user) The servers send the share of the results to the user who reconstructs the real result.[234]

3.1.1.3. Differential Privacy

Differential privacy is a statistical disclosure control technique using randomization by noise addition. Therefore, the outcome of differential privacy is never an exact but an approximation.[235] In this case it is used to provide broad access to summaries of sensitive data in a privacy preserving-way so that the output of the computation can be released to the general public.[236]

[230] source: Dan Bogdanov (n 226)
[231] Kamm and Willemson (n 228) 3
[232] ibid 3
[233] ibid 3
[234] ibid 4
[235] Alexandra Wood and others (n 54) 11
[236] ibid, 15

3.1.2. Legal evaluation and risk assessment

3.1.2.1. Data Processing activities

It is essential to match the actors and their data processing activities in real-life with the abstract data protection provisions.

In order to properly classify the data processing activities, these need to be broken down into phases and these phases of processing need to be evaluated on their own. For the sake of better understanding the following points will discuss the processing activities chronologically from the first to the last in three categories, processing activities relating to 1) data acquisition 2) data analysis and 3) data application.

3.1.2.1.1. Acquisition

Anonymisation can naturally only begin after a certain processing step, especially in the case of processing for secondary research purposes. This is a so-called subsequent anonymisation.

The first processing activity is necessarily the collection of data from the patients.[237] This typically happens face-to-face between the hospital personnel (a nurse, nurse practitioner, doctor, etc…) and the patient, which means that at this stage the data is for sure personal data. As such, the GDPR is fully applicable, first and foremost the provisions for the principals and lawfulness of processing.

The obtained personal data during the collection of the patient's account of their medical history about the patient's physical state, diabetes, health conditions all belong to the special categories of personal data pursuant to Art. 9 (1) GDPR, namely, as health data.[238] Datasets containing special categories of personal data clearly present a need for caution.

The first and foremost obligation of the hospitals as health care providers acting as data controllers is to lay down the purpose of the data processing and a corresponding appropriate legitimate basis in accordance with Art. 5, Art. 6 and Art. 9 respectively. Since in this case research is not the principal reason of the data collection, first the lawfulness of the original processing shall be assessed. As discussed above, the legal bases of processing for the purpose of treating the patient shall be assessed twofold. First, it is necessary that the processing can be assigned to one of the legal bases in Art. 6, which in this case would be Art. 6 (1) lit b), processing for contractual necessity. Let's assume that in this case the processing of the patient's general personal and health data is based on the treatment contract with the healthcare provider. It is lawful, since, as shown above, processing patient data for the purpose of treatment falls in the category of the exception in Art. 9 (2) lit. h).[239] Purpose and legal basis shall be determined at the time of the admission process, because that is when the active patient data processing begins within the framework of the hospital information system used in each case.[240]

With the signing of the treatment contract, the hospital is obliged to define and ensure all necessary measures for the protection of the patient's rights in all organisational units and during the entire patient's journey within the hospital. When drawing up the treatment contract, it is necessary to pay attention to certain points relevant from a data protection point of view. Although the content of the treatment contract is predominantly defined by the hospital, some questions have to be clarified individually with each patient. In this respect, an active involvement of the patient is also required when the treatment contract is concluded, so that the patient either makes certain statements or decides for or against a certain option (physician's letter, denomination, photo documentation, possible blocking notice).[241]

[237] cf. Peter Schantz, Art. 6 section 1 para. 122 in Simitis, Hornung and Spiecker (eds) (n 9)

[238] for the meaning of health data see 2.1.1.

[239] see section 2.2.2.2.2.

[240] Benedikt Buchner, *Datenschutz im Gesundheitswesen* (2nd edn. AOK Verlag 2019) ch C 204

[241] Ibid, 207

The central question regarding to the further use of already existing datasets of patient's records and medical data for secondary research purposes[242] the compatibility test is of pivotal importance. Based on the argumentation derived above, given that in this case the original legal basis of the processing was the treatment contract with a healthcare provider, there is no legal obstacles getting in the way of a compatibility assessment according to Art. 6 (4). This means that the compatibility-presumption of privileged secondary research purposes according to Art. 5 (1) lit b) in conjunction with Art. 89 (1) can unfold its full effect.[243] Consequently, anonymising patient's data for further research is a purpose considered not incompatible with the original purpose, and no legal basis other than the one used for the original processing is necessary.

3.1.2.1.2. Analysis

In many research projects, and especially those for statistical purposes, anonymised data processing is regularly sufficient to draw conclusions from a larger database or to show statistical significance after the individual personal data have been collected.[244] As far as the responsible party only wants to gain general knowledge from sensor data, it does not matter, who the concrete person behind the data is.[245] Besides, there is a strong incentive to anonymise health data, since this kind of personal data is highly sensitive and protected by strong, often limiting data protection provisions. In order to avoid privacy pitfalls, anonymisation should occur at the earliest possible opportunity, ideally prior to using the data for research purposes.[246]

The SODA model follows this approach by de-identifying personal data with help of secret sharing.[247] As discussed above, the anonymisation process itself is widely accepted to constitute the processing of personal data.[248] Since the input data for secret sharing is personal data in plaintext, this is not different in this case either. One counterargument found in legal literature is, that dividing data into illegible shards does not affect the substance of the data, it simply alters the form of the information the data holds by splitting it into smaller data chunks. According to this concept the alteration of the form of the data is not a processing activity within pursuant to Art. 4 Nr. 2 GDPR and consequently, it does not fall under the scope of data protection legislation.[249] This reasoning is difficult to reconcile with the broad definition of data processing though. The GDPR follows an approach that sees every and any treatment that '*happens to*' the data as data processing. As shown above, the fact that a given specific processing method or modality is not named in the list in Art. 4 Nr. 2 is not a sufficient argument. Firstly, because that list is not an exhaustive list, and secondly, because there are some blanket terms defined there, such as "*usage*" of personal data, that could be considered as fallback terms.[250]

Nonetheless, the fact that the data fragmentation procedure as such is processing of personal data does not mean that the output data has to fall under the scope of the GDPR. On the contrary, based on the arguments put forward here the data shards that have undergone the partitioning are considered to be non-personal data. The qualification of split data is relatively new to data protection law. There is, however, an analogy to borrow from intellectual property law. Even though there are naturally differences between data protection law and intellectual property law, one interesting parallel can be drawn. that if a copyright protected work is split in many parts, and those parts can be individually perceived, the copyright protection still applies for and protects the single parts. Supposing that those

[242] Bertram Raum, Art. 89 Rn. 18 in Ehmann and Selmayr (eds) (n 40)
[243] see section 2.2.3.
[244] Stephan Pötters, Art. 89 para. 12, in Gola (ed) (n 39)
[245] Peter Schantz, Art. 6 section 1 para. 122 in Simitis, Hornung and Spiecker (eds) (n 9)
[246] ICO (n 107) 46
[247] Bogdanov (n 226) 24 ff.
[248] see section 2.2.1.3.
[249] Tobias Herbst, Art. 4 Nr. 2 para. 25 in Kühling and Buchner (eds) (n 21)
[250] Marian Alexander Arning and Tobias Rothkegel, Art. 4 para. 82 in Taeger and Gabel (eds) (n 23); Tobias Herbst, Art. 4 Nr. 2 para. 28 in Kühling and Buchner (eds) (n 21)

individual parts of the copyright protected work do not allow the perception of the work as such without having all other parts of the work (such as a .zip file) the copyright protection might not apply for these parts.[251] Similarly, in case of data shards, the original data can only be restored and turned back into personal data if all fragments are put together.

Furthermore, it is not enough to state that in general, data fragmentation via secret sharing results in anonymised data, since the anonymous nature of the data is context related. Therefore, it must be assessed whether the data is anonymous in relation to the hospitals partaking in the research. As shown above, in the acquisition phase each hospital collects health data and they stay in possession of those medical records in plain form. This excludes the possibility of viewing their own data as anonymous data. Given the functioning of secret sharing and the differences between traditional encryption and secret sharing it is not likely that a single part of a secretly shared data 'package' enables the re-identification of a person. Firstly, the system itself offers protection against intruders on a technical level. Secondly, given the processing situation and aligning interests of the participating hospitals in the successful research the collusion of the parties is highly unlikely. These circumstances support the notion that the data keeps its identifiable and therefore personal nature in relation to the hospital where it was originally collected and is being stored, whereas it could be considered anonymous for the two other receiving institutions.

If personal data is turned into non-personal data, then the subsequent storage of the data pieces should not be considered 'data processing' within the frames of Art. 4 Nr. 2 and the data protection provisions in general.[252] Following this logic, the analytics carried out using MPC shall be considered as analysis carried out with feature data. In this case the MPC analytics software is available to the hospitals as software as a product (SaaP). SaaP requires the client – the hospitals participating in the research – to obtain a license for a software solution that will be hosted on the client's computer.

It is important to highlight the fact that no de-identification method can guarantee anonymised information without individual assessment. According to the definition of personal data and the absolute concept of identifiability with relative traits the GDPR seems to have opted for, several factors shall be assessed to determine the likelihood of occurrence of a re-identification. [253]

As to the technical details, structure and design of the processing, MPC is a state-of the-art privacy preserving tool. The cryptographic solutions in use protect the data from intruders during the analysis. Intruders in this sense mean unauthorised external adversaries not intended to have access to the anonymised data.[254]

To avoid the inadvertent data breaches, such as accidental destruction, alteration or unauthorised disclosure[255] further organisational measures shall accompany the implementation of MPC. These could be for example (electronic) access control, physical access control, entry logs, availability control.[256] Data protection related education and courses for the staff is also a feasible way to increase the level of security and protection.[257]

Apart from the technical and organisational measures a strong contractual framework can lower the risk and likelihood of unintended re-identification. For this reason, binding commitments or similar legal instruments aimed at preventing re-identification shall be included in any agreement for the processing

[251] Ernesto Damiani and others, in Stelvio Cimato (ed) '*D31.1 Risk assessment and current legal status on data protection*' [2014] deliverable produced within the frames of the PRACTICE project

[252] Hon, Millard and Walden (n 220) 11

[253] see section 2.2.1.2.

[254] Data Protection Commission of Ireland (n 107) 8

[255] Art. 4 Nr. 12 defines the term 'data breach' as a "*breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed*".

[256] Paul Voigt, *IT-Sicherheitsrecht,* (Otto Schmidt Verlag 2018) 128-129

[257] Sönke Maseberg, 'Technische und Organisatorische Maßnahmen' in Uwe Schläger and Jan-Christian Thode (eds.) *Handbuch Datenschutz und IT-Sicherheit* (Erich Schmidt Verlag 2018) ch I 1.3 530ff

of anonymised data.[258] Non-disclosure agreements may be another way to guarantee the impediment of identification, since such contractual clauses are legal prohibitions of recombining data from different providers.

3.1.2.1.3. Application

As shown above, under the circumstances of this case output research data can be considered anonymous data. According to the thorough evaluation provided in the opinion of the Art. 29 WP on anonymisation techniques, differential privacy is robust against all of the identified threats: singling out, linkability and inference. It preserves most of the data utility while protecting individual privacy. Hence it might be the best method to prevent identification and provide anonymous data.[259]

3.1.2.2. Involved Parties

Operationally speaking, there are three types of parties on the computational level:

1. Input Party: party which contributes input for the computation
2. Computing Party: party which takes part in the computation
3. Result/Output Party: party which receives the output of the data analysis.

For an accurate legal evaluation these parties need to be subsumed under one of the definitions the GDPR works with in Art. 4 for actors participating in or related to the data processing.

Starting with the classification of the input parties, the original data donors are the patients of the hospital. They are the natural persons to whom the identifiable information relates, in other words, the data subjects. They provide the hospitals with their personal data.

It is far more complicated to determine who the data controller is. The role of the controller is not fixed to one of the many participants. Resulting from the basic function of MPC every hospital here may play all the roles as input party, computing party and output party to cooperatively carry out the research and compute pre-defined functions. The role of the data controller is not fixed to one of the three participants, and that is perfectly common that more entities act as data controllers in relation to the same data processing. As outlined above, the two elements defining data controllers, the determination of means and purposes. In this case, the three hospitals commonly determine the what they want to do with their combined datasets, namely, statistical research with medical data of patients with diabetes for prediction of population health. Thus, the determination of purpose is given. Opting for MPC can be considered an autonomous conscious choice of the institutions as well, therefore, the determination of means is established as well. Since the hospitals decided on these factors together, eventually, they have to be considered as joint controllers pursuant to Art. 26 (1) GDPR, so that they are each responsible for actions to be taken.

Referring back to the processing activities, one should not forget that Art. 26 does not constitute an additional specific legal basis for processing by several controllers.[260] Rather, the processing by each individual controller requires an additional distinct legal basis as defined in Art. 6 (1). Those who determine the purpose and means of the processing must also ensure that the processing is carried out on an adequate legal basis respectively.

An important element of the situation outlined in this case it the use of private clouds by the participating hospitals. When an actor uses its own private cloud, it itself acts as cloud provider and cloud user simultaneously. In terms of data protection this could be translated into "in-house processing". This

[258] Data Protection Commission of Ireland (n 107) 8

[259] Art. 29 WP, WP 216 (n 107) 24

[260] Maria-Urania Dovas, 'Joint Controllership – Möglichkeiten oder Risiken der Datennutzung?' (2016) ZD 512, 516; Nikolaus Bertermann, Art. 26 para. 11 in Ehmann and Selmayr (eds) (n 40)

means that this actor making use of its own private cloud acts as data controller according to Art. 4 Nr. 7 GDPR.[261] The data controller does not only mean the institute as such, but also its employees. The persons who process personal data under its direct supervision are attributed to the controller under the premise that they are integrated into the organisational structure of the controller and are legally under its direct control. If this requirement is affirmed, it remains a processing operation of the controller and not an activity of a processor.[262]

3.2. Case 2: Intra-sector external MPC

The second case is an MPC analysis where the hospitals involved in the research want to share data of patients they have treated with diabetes. The data processing during the treatment was based on the explicit consent of the patients. The research would focus on prediction of population health, for instance to predict the risk of other comorbidities.
The hospitals perform joint analysis using data from different data providers in the same sector. The MPC software is run on the data using the resources of an external storage provider.

This case is a modification of the first one. Although the functioning is essentially the same there are several changes in the overall setting. These minor differences create new and different data protection related issues. These are:

1. lawfulness of the processing is based on consent pursuant to Art. 9 (2) lit. a);
2. outsourcing, data processing using of different levels of cloud services.

The circumstances of the analysis make this case more complex, with more factors that needed to be taken into account. The following appraisal focuses on the questions that differ from the ones of Case 1 along the same point.

3.2.1. Functioning

The architecture of the analysis is similar to that described in Case 1. MPC is combined with secret sharing and differential privacy. The difference lies within the design how the data processing activities are carried out. As opposed to Case 1, where the participating hospitals used their own private cloud, in this case the analysis is outsourced, and public cloud is used.

3.2.2. Legal evaluation and risk assessment

3.2.2.1. Data Processing activities

3.2.2.1.1. Acquisition

The collection of data occurs similarly as in the first case. None of the participating hospitals have collected the data primary for research purposes at the time of the original data acquisition. For the sake of giving a thorough evaluation it is assumed that in this case the patient’s consent was the chosen legitimate basis of the original data collection.

[261] cf. Georg Borges and Kirstin Brennscheidt, ’Rechtsfragen des Cloud Computing – ein Zwischenbericht’ in Georg Borges and Jörg Schwenk (eds) *Daten- und Identitätsschutz in Cloud Computing, E-Government und E-Commerce* (Springer, 2012) ch 3, 58
[262] Peter Schantz and Heinrich Amadeus Wolff, ‘*Das neue Datenschutzrecht, Datenschutz-Grundverordnung und Bundesdatenschutzgesetz in der Praxis*’(C.H.Beck 2017) ch 3 para 359

Given that medical data belongs to the special categories of data, the hospitals as data controllers must comply with the higher requirements of obtaining a valid consent according to Art. 9 (2) lit. a) GDPR. Subsequent anonymous processing does not remove the requirement for explicit consent.[263] This is to be understood to mean that the explicit consent must relate to the processing of sensitive data. The data used must be specifically named, as well as the processing purposes (in this case treatment). Also, as already mentioned above, an implied or tacit declaration should be excluded.[264] However, explicit does not mean in writing. It is only important that the data subject is sufficiently aware of the scope of the declaration, but this can also be adequately ensured by an electronic form. Oral consent, on the other hand, is not possible because of the necessary warning function and cannot be regarded as explicit consent within the meaning of Article 9 (2) lit. a).[265] As a result, it can be stated that consent must be obtained either in writing or electronically.

Data processing activities where the lawfulness of the processing is based on consent cannot be subject to a purpose compatibility assessment according to Art. 6 (4). Consequently, the processing cannot be continued further for another different purpose, even if and despite of the fact that the purpose of the further processing would be a privileged one covered by the presumption of Art. 5 (1) lit. b).[266] However, further processing on the basis of a different lawful basis for the purpose of subsequent anonymised processing is conceivable. This could be Art. 9 (2) lit. j) if there is a relevant Union or Member State legal basis. Otherwise, processing for statistical purposes may be based on consent. R 33 provides for simplified conditions for consent for such purposes. Under these conditions, consent can be given on a broader basis. This is really feasible in particular for Big Data analytics where the results and therefore indirectly the purposes cannot always be fully anticipated prior to the analysis.[267] In this way the decision is left to the data subjects as to what extent they want to sign a consent form with a broader wording for the purposes of the given research.[268] Consequently, in this case an informed, explicit consent is compliant if it contains a broadly formulated purpose such as *'anonymisation of your medical data with privacy-preserving technology for further research to improve population health*'.

3.2.2.1.2. Analysis

Up to the point of secret sharing the analysis occurs the same way as in the first case. The data stays identifiable and therefore personal data in relation to the hospital where it was originally collected and is being stored. It could be considered anonymous for the two other receiving institutions and for any other external third party.

The difference is that in this case the servers where the MPC analysis happens are not on-site. As it can be seen on Figure 6, the servers are hosted in the cloud in data centres operated by a cloud service provider, more precisely, an IaaS provider. IaaS provides computing resources such as processing power and/or storage for the processed data.[269] An IaaS provider is an external party who offers computing power with no involvement in the particular application.

In this case, MPC is provided to the participating parties in form of SaaS. This type of software is delivered online and is hosted by the software vendor or another external third party as cloud service provider.

IaaS is a low-level while SaaS is a high-level functionality. Pairing these two is a common combined service provision model. A software that costumers use and consider as SaaS might uses another cloud provider's IaaS services to perform their own.[270] This is precisely the service provision model of this case.

[263] Sebastian Schulz, Art. 9 para. 16 in Gola (ed) (n 39)

[264] see section 2.2.2.2.1.

[265] Britta Alexandra Mester, Art. 9 Rn. 19 in Taeger and Gabel (eds) (n 23)

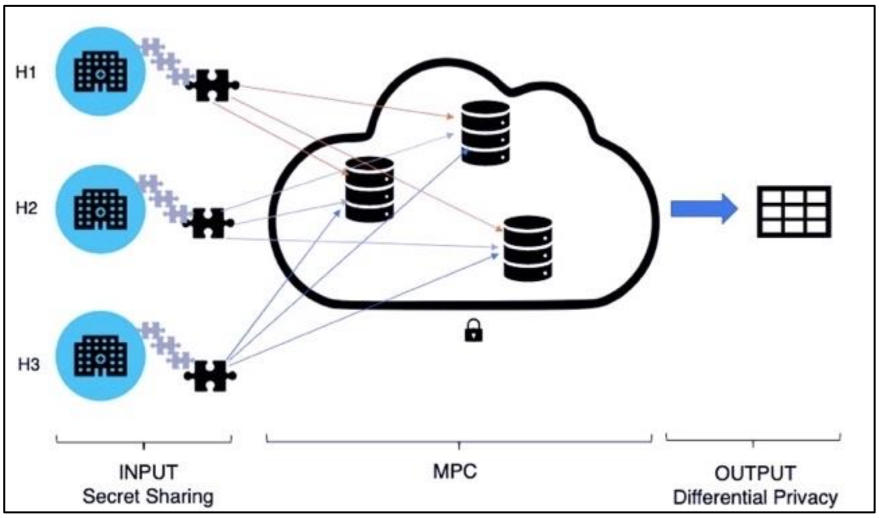
[266] see Figure 3 in section 2.2.3.

[267] R 33 s 2, 3

[268] Johannes Caspar, Art. 89 para. 37 in Simitis, Hornung and Spiecker (eds) (n 9)

[269] Hon, Millard and Walden (n 220) 4

[270] ibid, 5; Art. 29 WP, WP 196 (n 217) 5



6. Figure Case 2 setting

Although outsourcing the analysis has enormous benefits for the research, it comes with certain specific data protection risks. Within the frames of this case, the risks of this setting can be divided into two sub-categories. First, there are some inherent security risks in cloud computing in general, and second, there are specific risk factors when it comes to processing secret shared data in the cloud with MPC.

The risks in the first group can be traced back to the fact that clients do not have the same exclusive control of the data and the range of technical and organisational measures they are able to deploy might also be narrower.[271] Another issue is the lack of transparency, since what the cloud client does not know about, cannot assess as a data controller either. This problem typically arises in connection with information about the data processing chains (sub-contractors) or data transfers within and outside the European Economic Area.[272] Regardless, cloud clients acting as data controllers bear full responsibility for choosing reliable cloud service providers, who offer appropriate level of data security through satisfactory technical and organisational measures.[273]

A more immediate issue for MPC is how the use cloud-based processing, more precisely IaaS and SaaS, affects the anonymous nature of the data. In Case 2 the hospitals (cloud clients/users) would perform the anonymisation procedure before disclosing the resulting data in the cloud. The question is whether individual data shares are personal data for the cloud provider in a distributes analysis and storage setting on the provider's servers. As discussed in detail above, the secret shares of the data are not just a "re-organisation" of the data but a qualitative change, in content, where the identifiable character is removed, and the shares alone are non-intelligible. More specifically, they are only intelligible to the user.[274] One neuralgic point of this setting is that eventually, all the secret shared data fragments end up in the hand of the same cloud SaaS provider during the analysis. The protection against such intrusion or colluding is already provided by the system itself. The main advantage of a secret-shared database is its high level of confidentiality. It is nearly impossible for an individual computing party storing the data to learn anything about the values provided by the input parties. This is because of the security of the secret sharing scheme.[275] One of the properties MPC is supposed to ensure is input privacy, which means that the only information that can be inferred during the execution of the protocol is whatever

[271] Art. 29 WP, WP 196 (n 217) 2
[272] ibid, 6
[273] ibid, 14
[274] Hon, Millard and Walden (n 222) 221
[275] Bogdanov (n 226) 57

could be inferred from the output of the function. Another solution could be on a technical level that the operations on the data are distributed running simultaneously in different nods or locations. Afterwards these sub-operations would be combined and sent back to the client.[276] The decisive factor from the technical side is what level of security the access system implemented by the provider can guarantee.

On the legal side, the contractual framework surrounding the processing is of outstanding importance. Apart from the commercial contract signed by the cloud service provider and the cloud provider the data processing agreement signed by the parties acting as data controller and data processor respectively can specify the individual circumstances of the processing. Furthermore, a non-disclosure agreement alone or as an addendum lowers the risks of unauthorised re-identification, since that would mean a breach of contract. Liability clauses, remedies and contractual penalty could ensure the compliance with and fulfilment of the contract.

3.2.2.1.3. Application

The output of the MPC analysis is protected using differential privacy. As mentioned above, the Art. 29 WP considers differential privacy to be an efficient anonymisation tool. Naturally, whether data can keep its anonymous nature depends on the circumstances of the application. Both the form and the method of later data sharing can have an impact on the nature of the data.[277] Basically, the hospitals have two options in this case. Either they use the findings themselves or they make the results available for the wider scientific community.[278] Different kind of applications create different kind of risks of their own.

In the first scenario, the anonymous datasets are not disclosed to external third parties but are used only by the institutions who participated in the research. This means that the results can only be accessed by a well-defined, finite number of researchers. A major positive aspect of this situation is that the risks of re-identification are better controllable.[279] The hospitals would have the power and the obligation to put strong organisational measures in place in order to protect against unwanted re-identification, such as training the staff on security and arrangements for technical and organisational security (confidentiality agreements), controls over the ability to bring new data into the environment or limitation of use.[280]

These further restrictions are out of the hands of the original data controllers in case of publication and disclosure to the word at large. It cannot be guaranteed with the same certainty that data will be kept secure.[281] The risk exposure to linkability and inference is much higher in case of publication than it is in case of limited access.

3.2.2.2. Involved Parties

Like in the first case, the data subjects are the patients. The participating hospitals shall be considered joint controllers pursuant to Art. 26 (1) GDPR. Being a cloud client does not change this fact since ultimately, they set the purposes and decide on the outsourcing of the processing. Delegating part of the processing activities to external providers is a right of the data controller that does not affect their status.[282]

[276] ibid, 222

[277] cf. Art. 29 WP, WP 216 (n 107) 12

[278] There are several levels of disclosure between these two extremes ranging from restricted access to public availability without any access restriction. For an overview on this „spectrum for data access" see Khaled El Emam, *'Guide to the De-Identification of personal health information'* (CRC Press Taylor&Francis Group 2013) 3-4

[279] ICO (n 225) 37

[280] ibid, 37-38

[281] ibid, 37

[282] Art. 29 WP, WP 196 (n 217) 7

The interesting question here is the position of the cloud providers. As compared to private clouds, in case of a public cloud rollout model, the roles of cloud client and cloud provider are usually separated. Public cloud essentially means a cloud infrastructure owned by an organisation selling cloud services that is made available to a large group of users.[283] Normally, the cloud client, who uploads data to the cloud remains data controller and the cloud provider who processes that data on behalf of the cloud client (data controller) acts as a data processor pursuant to Art. 4 Nr. 8 GDPR.[284] It can therefore be said that if the analysis involved personal data, they would clearly be considered data processor.

This suggests that the status of the cloud service provider is dependent upon the nature of the data.[285] If and when the data is personal data, the position of cloud providers fulfils the requirements of being a data processor according to Art. 4 Nr. 8 in conjunction with Art. 28 GDPR. However, if data in the cloud could be deemed non-personal anonymous data, the GDPR would not apply anymore and they would not be data processors within the meaning of data protection laws. The main determinant regarding this question is, as shown above, the likelihood of identification. The possibility and probability of re-identification must be assessed based on the approach the GDPR opted for, which is a relative concept of identifiability, by taking into account objective factors.[286] This leads to the problem as to what are the "*means reasonably likely to be used to re-identify by the controller or any other person*".[287] 'Any other person' includes data processors as well. Whether means are reasonably likely to be used by the data processor depends on its motivation to re-identify and the available technological opportunities. If re-identification would be technologically possible, but the data processor is not motivated to do so, the use of those means can hardly be considered reasonably likely and vice versa. In this case there is a layered cloud service provision model where a SaaS provider uses the infrastructure of an IaaS provider. The test of reasonably likelihood must be carried out separately in relation to every data processor. From a technical point of view, given that the secret sharing happens prior the data is transmitted to the SaaS provider, this latter does not have access to the original dataset. Consequently, the IaaS provider, who offers its services to the SaaS provider, does not have access to the original dataset either. MPC is supposed to protect against inference during the computation, so that intruders cannot extract information about the plain data. From a legal point of view, a strong contractual framework including precise terms and additional incentives to adhere to those terms (such as contractual penalty) should ensure compliance. Arguably, assuming that the abovementioned technical and contractual safeguards are properly in place, cloud providers do not fall under the scope of the GDPR.[288]

3.3. Case 3: Inter-sector external MPC run by outsourcing

In the third case, an academic hospital, a public health-care insurance company and a research organisation aim to shorten the lifecycle between research and clinical practice by combining data from multiple verticals, thus improving healthcare efficiency. In particular, the case combines population data, patient profiles, patient care, and financial claim data. One application is comorbidity-based risk stratification for heart failure patients. The expectation is that combining clinical data from the hospital with cardiovascular comorbidity information based on claims data from the insurance company leads to better treatment outcomes. The results of the predictive analysis would be published and available for public use.

[283] ibid, 25

[284] ibid, 8

[285] critical Hon, Millard and Walden (n 220) 11

[286] see section 2.2.1.2.

[287] R 26 GDPR,

[288] Hon, Millard and Walden (n 222) 219

This case is a second modification of the Case 1. It includes the more complex processing setting of Case 2. This time the difference is not in the technical and organisational details, but in the initial situation. This time the differences are that the participating organisations

1. come from different branches,
2. contribute to the research with different kind of personal data and
3. make use of different provisions of the GDPR regarding to the lawfulness of processing.

3.3.1. Functioning

The architecture of the analysis is similar to that described in Case 1 and Case 2. MPC is combined with secret sharing and differential privacy. Similar to Case 2, the analysis is outsourced, and cloud computing services are used.

3.3.2. Legal evaluation and risk assessment

This case presents the medical sector as an interconnected system of medical service providers and other actors. The cooperation between hospitals, health insurance companies and research institutions establishes an efficient system of health care and health research. These actors can only fulfil their respective or common tasks if they have the necessary information at their disposal. Data exchange is therefore indispensable.[289]

3.3.2.1. Data Processing activities

3.3.2.1.1. Acquisition

The collection of data occurs similarly as in Case 1 and Case 2, originally not necessarily with the intention of carrying out future research. However, in this case the different parties had different primary purposes.

The first institution is an academic hospital. Just like in Case 1 and 2, the original purpose of the very first data collection in the academic hospital was the treatment and care of the patients. Let's assume that similarly to Case 1, the lawfulness of the processing for this purpose was based on a treatment contract in accordance with Art. 9 (2) lit. h) GDPR, that is, processing necessary for the purposes of preventative medicine, on the basis of contract with a health professional.

The second involved institution is a public health insurance company. Health insurance companies process a wide variety of personal data. Traditionally, insurance companies process common demographic data (age, marital status, profession, address, phone number), population data usually in an aggregated form (e.g. mortality and morbidity rates), medical data or health-related behavioural data (e.g. smoking or drinking habits). Thanks to digitalisation new sources of data have emerged, and therefore insurance companies might get hold of online medical data, IoT data or financial data (e.g. bank account data or credit card data).[290] Not all of these data belong to the special categories of personal data as defined in Art. 9 GDPR. The processing of regular personal data, such as certain demographic data or financial claim data itself, can be based on one of the lawful bases listed in Art. 6 (1). Depending on the specific circumstances the appropriate provision could be the data subject's consent according

[289] Anne Paschke, 'Datenschutz im Medizinsektor' para. 65 in Louisa Specht and Reto Mantz (eds) (n 142)

[290] European Insurance and Occupational Pensions Authority, '*Big Data Analytics in Motor and Health Insurance: A Thematic Review*' (April 2019) 9 f, available at https://eiopa.europa.eu/Publications/EIOPA_BigDataAnalytics_ThematicReview_April2019.pdf accessed 22 December 2019

to Art. 6 (1) lit. a), contractual necessity according to Art. 6 (1) lit. b) or the data controller's legitimate interest as regulated in Art. 6 (1) lit. f). Data that is highly sensitive and falls under Art. 9, such as medical data or health related behavioural data shall be processed based on one of the exemptions provided by Art. 9 (2). The appropriate provision is typically Art. 9 (2) lit. h), which allows the processing if it is necessary for the management of health and social care systems and services on the basis of Union or Member State law.[291] Management of systems in the health or social field is aimed in particular at the providers of the costs of health or social services, i.e. the health insurance funds. This follows from Recital 52 s 2 as well, which clarifies that the economic viability of health insurance systems is also a public interest objective that may justify an exception to the prohibition of processing special categories of data.[292]

The third participating institution is a research institution. This is the first example where an involved actor could actually collect data directly for primary research purposes. In this case, the lawfulness of the processing shall be based on Art. 9 (2) lit j) GDPR. This provision provides an exemption from the general prohibition of processing special categories of personal data by an opening clause. It allows the processing of health data if this processing is "*necessary for […] statistical purposes in accordance with Article 89 (1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject*".[293] As opposed to subsequent secondary research, here there is no need for a compatibility presumption, since the legitimate basis of the processing directly allows for the processing of statistical research purposes.

3.3.2.1.2. Analysis

The analysis in this case is carried out essentially in the same way as in Case 2.

3.3.2.1.3. Application

Case 3 addresses a statistical research that was carried out with the intent of publishing its results and findings. As mentioned above, in case of disclosing anonymous data to a large community where the circle of recipients is not previously defined, the risks of re-identification are generally higher.[294] If there are no restrictions attached to an open access, nothing prevents the recipients from analysing the data further in any way they deem useful. This may entail linking the dataset with their own data or combining it with other datasets and making these new datasets publicly available once again.[295]

Privacy preserving tools are one way to manage the relatively high risks of publicly available datasets. If the risk of disclosure is high, data should be aggregated to a level where these risks are kept at the minimum, taking into account granularity, the size of the dataset or the estimated circle of recipients.[296] If the transaction risks – that is, the risks of disclosure – are high, strong and robust privacy preserving methods must be used, regardless the additional costs. Operating otherwise could lead to unfavourable legal, regulatory and, ultimately, financial consequences.[297]

Against this background it can be said that the most serious mistake of differential privacy in terms of data protection is to not generate enough noise to add to the data. If too little noise has been injected, then even if the data as such is indeed useful, the privacy of the individuals is not protected.[298] This

[291] Thilo Weichert, Art. 9 para. 106 ff in Kühling and Buchner (eds) (n 21);
[292] David Kampert, Art. 9 para. 45 in Ehmann and Selmayr (eds) (n 40)
[293] the relevant provision in national law being in Germany being § 27 FDPA
[294] see section 3.2.2.1.3.
[295] El Emam (n 278) 3
[296] ICO (n 225) 36
[297] El Emam (n 278) 5 ff, who calls this a „*balanced risk management approach*"
[298] Art. 29 WP, WP 216 (n 107) 16, 28

would essentially mean a failed estimation of the privacy-utility trade-off. On the other hand, if applied correctly, differential privacy could be a feasible way to protect the individual's privacy and to retain the usefulness of the data at the same time.[299]

3.3.2.2. Involved Parties

In this case there are three clearly distinguishable circle of data subjects belonging to the three different participating actors.

	1.	**2.**	**3.**
data subject	patients	insured clients	research participants
data controller	academic hospital	public insurance company	research institution

The data subjects are who the data originates from, who the personal data relates to. In case of a hospital, they are the patients. At an insurance company, they are the insured clients who the insurance company signed insurance contracts with. A research institution can – like in this particular case – contact people directly via various ways where these people can voluntarily disclose necessary information.

Assuming that the academic hospital, the insurance company and the research institutions determine the purpose and at least the non-delegable means together, they should be considered joint controllers pursuant to Art. 26 GDPR. The fact that they provide different type of personal data as input data from different groups of data subjects does not affect this classification. Choosing the same lawful basis for the original data collection is not a prerequisite for joint controllership either. What is important though, is that the parties must lay down their respective responsibilities in an agreement pursuant to Art. 26 (1) and (2).

As regards to data processors, similarly to Case 2, in case proper technical and organisational measures as well as a strong contractual framework between the data controllers and between the data controllers as cloud clients and the cloud service provider accompany the application of MPC, the cloud providers could be removed from under the scope of data protection regulations.[300]

[299] ibid, 28

[300] see section 3.2.2.2.

4. Conclusion and closing remarks

The two aims of the SODA project

- to develop a scalable data analytics tool that does not lead to low data quality and
- is secure and suitable for preserving the privacy of the individuals at the same time

resonates well with the two-fold purpose of the GDPR and, ultimately, of the overall data ecosystem of the Digital Single Market of the EU. Choosing healthcare as demonstrator is worthwhile, since utilisation of Big Data in terms of health care could be extremely beneficial. Since health-related data is taken as one of the most confidential and personal kind of data, preserving the confidentiality and establishing the secure processing of such data is of utmost importance. For this reason, there is a high incentive for anonymised processing of health data.

The European mindset of data protection can best be thought of as a right of informational self-determination. Safeguarding this right in today's information society is a key issue in the current framework of the European data protection law. Resulting from the fact that data protection roots deeply from the fundamental rights framework, the definition of personal data is very broad and the level of protection by legal and regulatory instruments regarding this data is very high.

As discussed, the nature of the data is defined following a binary concept. Consequently, the leeway for data anonymisation and anonymous data directly depends on the concept of personal data. The most important feature of personal data from the perspective of anonymisation is identifiability. This threshold condition determines whether data has a personal relevance or can be considered anonymous data. Despite its practical relevance, currently there are no exact regulations or metrics to evaluate the risks of re-identification in advance. The two concurring approaches to this question are the absolute and relative concept of identifiability and it is difficult to arrive to a conclusion as to which of the two concepts is prevailing. An absolute approach would extend the scope of the GDPR essentially without any real or considerable boundaries. A relative approach, on the other hand, is more pragmatic and considers a reconciliation of interest of different stakeholders, accordingly, it simply imposes limitations on the scope of the data protection law, leaving the essence of data protection intact. Following a relative concept does not result in a protection loophole, since it does not change the protected purpose, the right of informational self-determination and integrity of the data subject. This would not contradict the purposes of the GDPR, since in scenarios where no realistic or reasonable chances to identify the data subject exist – in other words, the residual risk of identification is minimal – with respect to the processing in question, the protection offered by the GDPR is not affected at all. The GDPR itself does not settle this issue, however, there is a noticeable tendency of leaning on elements relativizing a strict absolute approach. With the reference for 'another person' in Recital 26 the GDPR categorically rejected a one-sided relative concept, which would entirely exclude the relevance of the potential knowledge of a third party. On the other hand, by including means reasonably likely to be used by the motivated parties, it acknowledges the importance of the particular context and the unique circumstances of a specific case.

Additional to the issue of identifiability this analysis identified three other challenges that inevitably occur in connection with the use cases. Firstly, when determining the lawfulness of the processing, the legitimate ground of the processing must be chosen in conjunction with the special provisions that apply for special categories of personal data. Secondly, when repurposing the original data processing operation for subsequent research, due account should be given to the complex interplay of Art. 5 (1) lit. b) on purpose limitation and the presumption of compatible purposes, Art. 6 (4) on compatibility assessment and Art. 89 (1) on processing for scientific research and statistical research purposes. Last but not least, the specific design of the data analysis must be assessed, since the relation of the participating parties to each other and to the original and de-identified datasets may influence the nature of the data during the data processing.

The three selected use cases address these problems from different perspectives. The first case starts with a relatively simple data processing design, with three institutions from the same branch, making use of their own resources. The second case builds upon the first, with the modification that the data analysis is outsourced, and the parties use cloud computing resources. The third case includes aspects of the two previous cases and examines them in an inter-sector collaborative data processing setting.

Essentially, by analysing similar cases it can be shown that even minor modifications in the processing design lead to different legal implications. Processing design should be understood broadly, and it does not only mean the technical details, but also the lawfulness of the entire processing. GDPR compliant data processing requires the implementation of appropriate technical measures, and an efficient contractual framework guiding it. The motivation to anonymise personal does not change this obligation, since generally the data is personal at the beginning of the processing activity.

The solution approach this study suggests is dual. On a conceptual level, the dual concept of anonymisation should be considered. Separating anonymisation as processing of personal data and anonymity as state of the data might help in clarifying the obligations of the parties. Anonymisation as processing falls under the GDPR, therefore, all the obligations related to the processing of personal data apply. This includes the principles and lawfulness of processing, obligations of controller and processor as well as data security provisions. Anonymous data, on the other hand, falls outside the scope of the GDPR with the reservation that there are no means reasonably likely to be used available to reverse the anonymisation. On a practical level, assuming that the relative concept of anonymity provides some leeway and personal data can be considered anonymous if there is no personal reference anymore, a context-specific risk assessment needs to be carried out at the onset, and regular review, continuous evaluation and comprehensible documentation is needed in order to make sure that the used technical and organisational measures are state-of-the-art and can protect the anonymised data.

The broad implication of the present assessment of the three use cases is that legally compliant data processing can be achieved through the structured implementation of technical and organisational measures as well as contractual safeguards. Cryptographic solutions such as multi-party computation have the potential to fulfil the requirements for computational anonymity by creating anonymised data in a way that does not allow the data subjects to be identified with means reasonably likely to be used.

The study concludes by arguing that Big Data and privacy do not need to be mutually exclusive. Finding a delicate balance between data protection and data utilisation is the key to the compliant handling of personal data.

List of Cases

European Court of Justice

Case C-300/95 *Commission of the European Communities v United Kingdom of Great Britain and Northern Ireland* [1997] ECLI:EU:C:1997:255

Case C-70/10 *Scarlet Extended SA v Sabam* [2011] ECLI:EU:C:2011:771

Case C-131/12 *Google Spain SL, Google Inc. v Agencia Española de Protección ded Datos (AEDP), Mario Costeja González* [2014] ECLI:EU:C:2014:317

Case C – 582/14 *Patrick Breyer v Bundesrepublik Deutschland* [2016] ECLI:EU:C:2016:779

Case C-434/16 *Peter Nowak v Data Protection Commissioner* [2017] ECLI:EU:C:2017:994

Case C-210/16 *Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig-Holstein GmbH* [2018] ECLI:EU:C:2018:388

Case C-25/17 *Korkein hallinto-oikeus v Tietosuojavaltuutettu* [2018] ECLI:EU:C:2018:551

Case C-40/17 *Fashion ID GmbH & Co. KG v Verbraucherzentrale NRW* [2019] ECLI:EU:C:2019:629

German Federal Court of Justice (BGH)

BGH VI ZR 135/13 *EuGH-Vorlage zur Speicherung von dynamischen IP-Adressen – IP-Adressen* [2014]

Bibliography

EDPB, Opinion 3/2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection Regulation (GDPR) (art. 70.1.b.)) (2019) available at https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_opinionctrq_a_final_en.pdf

EDPB, Letter to ICANN (2018) available at https://edpb.europa.eu/sites/edpb/files/files/file1/icann_letter_en.pdf

Art. 29 WP, Guidelines on transparency under Regulation 2016/679, WP 260 (2018) available at https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=622227

Art. 29 WP, Guidelines on Consent under Regulation 2016/679, WP 259 (2018) available at https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051

Art. 29 WP, Statement of the WP29 on encryption and their impact on the protection of individuals with regard to the processing of their personal data in the EU (2018) available at http://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=51026

Art. 29 WP, 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, WP 248 rev. 01 (2017) available at https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

Art. 29 WP, Guidelines on Data Protection Officer 16/EN WP 243rev.01 (2017) available at https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612048

Art. 29 WP, Statement of the WP29 on the impact of the development of big data on the protection of individuals with regard to the processing of their personal data in the EU, WP 221 (2014) available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp221_en.pdf

Art. 29 WP, Statement on the role of risk-based approach in data protection legal frameworks, WP 218 (2014) available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp218_en.pdf

Art. 29 WP, Opinion 05/2014 on Anonymisation Techniques, WP 216 (2014) https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

Art. 29 WP, Opinion 03/2013 on purpose limitation, WP 203 (2013) available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

Art. 29 WP, Opinion 05/2012 on Cloud Computing, WP 196 (2012) available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2012/wp196_en.pdf

Art. 29 WP, Advice paper on special categories of data (2011) available at https://www.pdpjournals.com/docs/88417.pdf

Art. 29 WP, Opinion 01/2010 on the concepts of "controller" and "processor", WP 169 (2010) available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2010/wp169_en.pdf

Art. 29 WP, Opinion 04/2007 on the concept of personal data, WP 136 (2007) available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf

Art. 29 WP, Health data in apps and devices, Annex to the response of the Art. 29 WP to DG Connect (Mr. Timmers) available at https://ec.europa.eu/justice/article-29/documentation/other-document/files/2015/20150205_letter_art29wp_ec_health_data_after_plenary_annex_en.pdf

Albrecht, J.P.; **Jotzo**, F, *Das neue Datenschutzrecht der EU* (Nomos 2017)

Auer-Reinsdorff, A**; Conrad,** I (eds) *Handbuch IT- und Datenschutzrecht* (3th edn C.H.Beck 2019)

BayLDA, Tätigkeitsbericht zur Sicherheit der Verarbeitung – Art. 32 DSGVO (2016) available at https://www.lda.bayern.de/media/baylda_ds-gvo_1_security.pdf

BayLDA, 8. Tätigkeitsbericht des Bayerischen Landesamts für Datenschutzaufsicht für die Jahre 2017 und 2018 (2019) available at https://www.lda.bayern.de/media/baylda_report_08.pdf

Bogdanov, D, *Sharemind: programmable secure computations with practical applications* (University of Tartu Press 2013)

Borges, G; **Schwenk,** J (eds) 'Daten- und Identitätsschutz in Cloud Computing, E-Government und E-Commerce' (Springer, 2012)

Brink, S; **Eckhardt**, J, *Wann ist ein Datum ein personenbezogenes Datum? Anwendungsbereich des Datenschutzrechts* (ZD 2015, 205-212)

Brisch, Klaus; **Pieper** Fritz, *Das Kriterium der "Bestimmbarkeit" bei Big Data-Analyseverfahren* 2015 CR 724

Buchner, B, *Datenschutz im Gesundheitswesen* (2nd edn, AOK Verlag 2019)

Burgin, M, *Theory of Information. Fundamentality, Diversity and Unification* (World Scientific Publishing 2010)

Bygrave, L. A., 'Information Concepts in Law: Generic Dreams and Definitional Daylight' (Oxford Journal of Legal Studies 2015 Vol. 35 No. 1 91-120)

Cavoukian, A, *Privacy by Design: From rhetoric to reality* (Information and Privacy Ontario, 2012)

Cavoukian, A, *Privacy by design – The 7 Foundational Principles, Implementation and Mapping of Fair Information Practices* (2012) available at https://www.iab.org/wp-content/IAB-uploads/2011/03/fred_carter.pdf

CIPL, The Role of Risk Management in Data Protection (White Paper, Cm, 2014) available at https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/white_paper_2-the_role_of_risk_management_in_data_protection-c.pdf

CIPL, Risk, High Risk, Risk Assessments and Data Protection Impact Assessments under the GDPR (White Paper, Cm, 2016) available at www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_gdpr_project_risk_white_paper_21_december_2016.pdf

CNIL: Methodology for Privacy Risk Management – How to implement the Data Protection Act, (2012) English language version available at https://www.cnil.fr/sites/default/files/typo/document/CNIL-ManagingPrivacyRisks-Methodology.pdf

Data Protection Commission of Ireland, 'Guidance Note: Guidance on Anonymisation and Pseudonymisation' (2019) available at https://www.dataprotection.ie/sites/default/files/uploads/2019-06/190614%20Anonymisation%20and%20Pseudonymisation.pdf

Datenschutzkonferenz, 'Datenschutz-Folgenabschätzung nach Art. 35 DS-GVO' (2018) 5 DSK-Kurzpapier, available at https://www.datenschutzkonferenz-online.de/media/kp/dsk_kpnr_5.pdf

Dettmeyer, R, *Medizin&Recht für Ärzte, Grundlagen – Fallbeispiele, Medizinrechtliche Fragen* (Springer, 2001)

Dovas, M-U, *Joint Controllership – Möglichkeiten oder Risiken der Datennutzung?* (ZD 2016, 512-517)

Ehmann, E; **Selmayr**, M, *Datenschutz-Grundverordnung – Kommentar* (2nd edn, C.H. Beck 2018)

El Khoury, A, '*Dynamic IP Addresses Can Be Personal Data, Sometimes. A Story of Binary Relations and Schrödinger's Cat'* (European Journal of Risk Regulation 2017 1, 191-197)

El Emam, K, *'Guide to the De-Identification of personal health information'* (CRC Press Taylor&Francis Group 2013)

ENISA, Working paper on 'Privacy and Data Protection by Design – From policy to engineering' (2015) available at https://www.enisa.europa.eu/publications/privacy-and-data-protection-by-design

Ernst, S, *Die Einwilligung nach der Datenschutzgrundverordnung – Anmerkungen zur Definition nach Art. 4 Nr. 11 DS-GVO* (ZD 2017, 110-114)

Esayas, S, *'The Role of Anonymisation and Pseudonymisation Under the EU Data Privacy Rules: Beyond the 'All or nothing' Approach'* (European Journal of Law and Technology 2015 6 No. 2) available at SSRN https://ssrn.com/abstract=2746831

European Insurance and Occupational Pensions Authority, 'Big Data Analytics in Motor and Health Insurance: A Thematic Review' (April 2019) 9 f, available at https://eiopa.europa.eu/Publications/EIOPA_BigDataAnalytics_ThematicReview_April2019.pdf

Friedewald, M; **Obersteller**, H; **Nebel**, M; **Bieker**, F; **Rost**, M, *Datenschutz-Folgenabschätzung – Ein Werkzeug für einen besseren Datenschutz* (White Paper); 2016. Available at: https://www.forum-privatheit.de/wp-content/uploads/Forum_Privatheit_White_Paper_Datenschutz-Folgenabschaetzung_2016-1.pdf

Gola, P, *DS-GVO Datenschutz-Grundverordnung – Kommentar* (2nd edn, C.H.Beck 2018)

Gierschmann, S; **Schlender**, K; **Stentzel**, R; **Veil**, W, *Kommentar Datenschutz-Grundverordnung* (1st edn, Bundesanzeiger Verlag 2018)

Forgó, N, *My health data—your research: some preliminary thoughts on different values in the General Data Protection Regulation* (IDPL, 2015, Vol. 5, No. 1, 54-63)

Härting, N, *Datenschutz-Grundverordnung – Das neue Datenschutzrecht in der betrieblichen Praxis* (1st edn, Verlag Dr. Otto Schmidt 2016)

Härting, N; **Gössling**, P, *Gemeinsame Verantwortlichkeit bei einer Facebook-Fanpage* (NJW 2018, 2523-2526)

Hintze, M; **El Emam**, K, *'Privacy Analytics White Paper – Comparing the Benefits of Pseudonymisation and Anonymisation Under the GDPR*' (*Iapp blog*, 17 August 2017) available at https://iapp.org/media/pdf/resource_center/PA_WP2-Anonymous-pseudonymous-comparison.pdf

Hon, K; **Millard**, C; **Walden**, I, *Who is responsible for 'personal data' in cloud computing? – The cloud of unknowing, Part 1* (IDPL, 2011, Vol. 1 No. 4, 211-228)

Hon, K; **Millard**, C; **Walden**, I, *Who is responsible for 'personal data' in cloud computing? – The cloud of unknowing, Part 2* (IDPL, 2012, Vol. 2 No. 1, 3-18)

ICO, Guidance on Data Protection Impact Assessments available at https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/data-protection-impact-assessments-dpias/when-do-we-need-to-do-a-dpia/

ICO, 'What about anonymised data?' available at https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/what-is-personal-data/#pd5

ICO, 'Anonymisation: managing data protection risk – code of practice' (2012) available at https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf

ICO, 'Special category data' (ICO Guide to the General Data Protection Regulation) available at https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/special-category-data/

ICO, 'When is consent appropriate?' available at https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/consent/when-is-consent-appropriate/

Jain, A, *The 5V's of big data* (2016), available at https://www.ibm.com/blogs/watson-health/the-5-vs-of-big-data/

Jandt, S, *Smart Health. Wird die DSGVO den dynamischen Herausforderungen gerecht?* (DuD 2016, 571-574)

Jäschke, T, *Datenschutz und Informationssicherheit im Gesundheitswesen* (MWV 2018)

Kamm, L; **Willemson**, J, *Secure Floating-Point Arithmetic and Private Satellite Collision Analysis* (International Journal of Information Security 3 2015, 14/6)

Kelleher, D, '*In Breyer decision today, Europe's highest court rules on definition of personal data*' (Iapp Blog 2016) available at https://iapp.org/news/a/in-breyer-decision-today-europes-highest-court-rules-on-definition-of-personal-data/

Knopp, M, *'Pseudonym – Grauzone zwischen Anonymisierung und Personenbezug'* (DuD 2015 39, 527-530)

Koops, BJ, *'The trouble with European Data Protection Law* (International Data Privacy Law 2014 4 (4) 250-265)

Kühling, J; **Buchner**, B, *Datenschutz-Grundverordnung – Kommentar* (2nd edn, C.H. Beck 2018)

Kühling, J; **Martini**, M, *Die Datenschutz-Grundverordnung: Revolution oder Evolution im europäischen und deutschen Datenschutzrecht?* (Europäische Zeitung für Wirtschaftsrecht 2016, 448-454)

Kühnl, C and others, '*Ein europäischer Gesundheitsdatenschutz*' (DuD 2018, 735-740)

Lord, N, 'Data Protection: Data in Transit vs. Data At Rest' (*Data Insider – The Guardian's digital blog*, 2019) available at https://digitalguardian.com/blog/data-protection-data-in-transit-vs-data-at-rest

Marnau, N, *'Anonymisierung, Pseudonymisierung und Transparenz für Big Data'* (DuD 2016 40, 428-433)

Monreal, M, *Weiterverarbeitung nach einer Zweckänderung in der DS-GVO* (ZD 2016, 507-512)

Moos, F, *Update Datenschutz* (DSRITB 2018, 259-273)

OECD, Supplementary Explanatory Memorandum to OECD Guidelines C(80)58/FINAL, 24, available at http://www.oecd.org/sti/ieconomy/2013-oecd-privacy-guidelines.pdf

Ohm, P, 'Broken promises of privacy: Responding to the Surprising Failure of Anonymisation' (UCLA Law Review 2010 Vol. 57 1701-1777)

Oostveen, M, *Identifiability and the Applicability of Data Protection to Big Data* (International Data Privacy Law 2016, Vol. 6/4 299-309)

Orlandi, C, *Is Multiparty Computation Any Good in Practice?* Conference: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2011 Prague Czech Republic 22-27 May 2011

Paal, B; **Pauly,** D.A, *Datenschutz-Grundverordnung – Kommentar* (3rd edn, C.H. Beck 2018)

Paterson, M; **McDonagh**, M, *Data Protection in an Era of Big Data: The Challenges Posed by Big Personal Data* (Monash University Law Review 2018, Vol 44, 1-32)

Plath, K-U, *Kommentar zum BDSG und zur DSGVO sowie den Datenschutzbestimmungen des TMG und TKG* (2nd edn, Verlag Dr. Otto Schmidt 2016)

Putrova, N. *'The law of everything. Broad concept of personal data and future of EU data protection law*' (Innovation and Technology 2018, 10:1, 40-81)

Richards, N.M; **King**, J.H, *Three Paradoxes of Big Data* (Stanford Law Review Online,Vol. 66, 41-46)

Rubinstein, I.S, *Big Data: The End of Privacy or a New Beginning?* (IDPL 2013, Vol. 3 No. 2 74-87)

Schaar, K, *Anpassung von Einwilligungserklärungen für wissenschaftliche Forschungsprojekte – die informierte Einwilligung nach der DS-GVO und den Ethikrichtlinien* (ZD 2017, 213-220)

Schantz, P; **Wolff**, H.A. '*Das neue Datenschutzrecht, Datenschutz-Grundverordnung und Bundesdatenschutzgesetz in der Praxis*'(C.H.Beck, 2017)

Schläger, U; **Thode**, J (eds.) *Handbuch Datenschutz und IT-Sicherheit* (Erich Schmidt Verlag 2018)

Siebel, M, *Abgrenzung der "allgemein anerkannten Regeln der Technik" vom "Stand der Technik"* (NJW 2013, 3003-3004)

Simitis, S; **Hornung**, G; **Spiecker**, I, *Datenschutzrecht DSGVO mit BDSG* (1st edn, Nomos 2019)

Specht, L; **Mantz**, R (eds) *Handbuch Europäisches und deutsches Datenschutzrecht* (C.H.Beck 2019)

Spindler, G, *Persönlichkeitsschutz im Internet - Anforderungen und Grenzen einer Regulierung* (Verhandlungen des 69. Deutschen Juristentages 2012 Vol I F 116)

Spindler, G, *'Die neue EU-Datenschutz-Grundverordnung'* (Der Betrieb 2016, 937-947)

Spindler, G, *Big Data und Forschung mit Gesundheitsdaten in der gesetzlichen Krankenversicherung* (Medizinrecht 2016, 691-699)

Spindler, G; **Schuster,** F, *Recht der elektronischen Medien* (4th edn, C.H.Beck 2019)

Spindler, G; **Schmechel,** P, *Personal Data and Encryption in the European General Data Protection Regulation* (JIPITEC 2016, 163)

Staunton, C; **Slokenberga**, S; **Mascalzoni**, D, *The GDPR and the research exemption: considerations on the necessary safeguards for research biobanks* (European Journal of Human Genetics 2019, 27, 1159-1167, available at https://www.nature.com/articles/s41431-019-0386-5)

Sydow, G, *Europäische Datenschutzgrundverordnung* (1st edn, Nomos 2017)

Taeger, J; **Gabel**, D, *Kommentar DSGVO – BDSG* (3rd edn, C.H.Beck 2019)

Van Alsenoy, Brendan, *Liability under EU Data Protection Law: From Directive 95/46 to the General Data Protection Regulation* (JIPITEC 2016 271)

Veil, W, *DS-GVO: Risikobasierter Ansatz statt rigides Verbotsprinzip – Eine erste Bestandsaufnahme* (ZD 2015, 347-353)

Voigt, P, *IT-Sicherheitsrecht*, (Otto Schmidt Verlag 2018)

Waen, H; **Van Essen**, J; **Wellens**, V, '*Confidentiality agreements are not data processing agreements*' (Lexology, 2015) available at https://www.lexology.com/library/detail.aspx?g=e1d5ccfb-f0a0-4c32-aa59-a65864af1acd

Wagner, B, *Disruption der Verantwortlichkeit, private Nutzer als datenschutzrechtliche Verantwortliche im Internet of things* (ZD 2018, 307-312)

Webber, M, *The GDPR's impact on the cloud service provider as a processor* (Privacy & Data Protection 2017 Vol. 16/4)

Werkmeister, C; **Brandt**, E, *Datenschutzrechtliche Herausforderungen für Big Data* (CR 2016, 233-238)

Wood, A; **Altman**, M; **Bembenek**, A; **Bun**, M; **Gaboardi**, M; **Honaker**, J; **Nissim**, K; **OBrien**, D. R; **Steinke**, T; **Vadhan**, S, *Differential privacy: A primer for a non-technical audience* (Vanderbilt Journal of Entertainment & Technology Law 21, no. 1 (2018) 209-275)

Wulf, H. M; **Burgenmeister**, C: *Industrie 4.0 in der Logistik – Rechtliche Hürden beim Einsatz neuer Vernetzungs-Technologien* (CR 2015, 404-412)

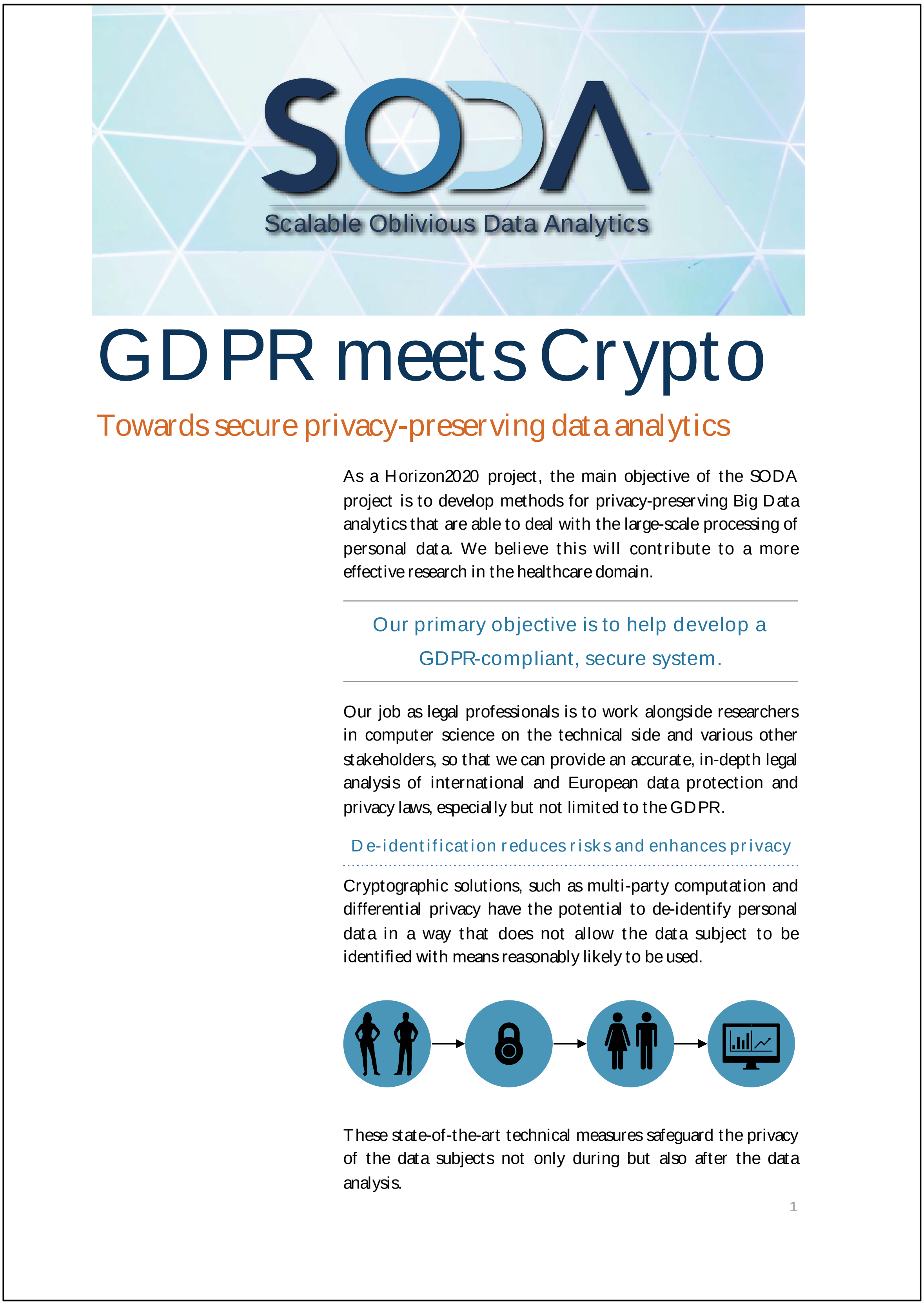
Zarsky, T. Z, *Incompatible: The GDPR in the Age of Big Data* (Seton Hall Law Review, Vol. 47/4 Article 2)

A. Appendix: SODA Brief

A brief overview of the core findings was used for promoting the exchange of expertise. The document was sent out to various data protection authorities and data protection officers to facilitate interaction between academia and practice and in order to gain a rather vocational perspective as well.

List of contacted DPAs and other stakeholders:

1. Data Protection Officer of Lower-Saxony, Germany
2. Independent Centre for Data Protection Schleswig-Holstein, Germany
3. Bavarian State Office for Data Protection Supervision
4. TNO, Netherlands Organisation for Applied Scientific Research



SODA

Scalable Oblivious Data Analytics

GDPR meets Crypto

Towards secure privacy-preserving data analytics

As a Horizon2020 project, the main objective of the SODA project is to develop methods for privacy-preserving Big Data analytics that are able to deal with the large-scale processing of personal data. We believe this will contribute to a more effective research in the healthcare domain.

> Our primary objective is to help develop a GDPR-compliant, secure system.

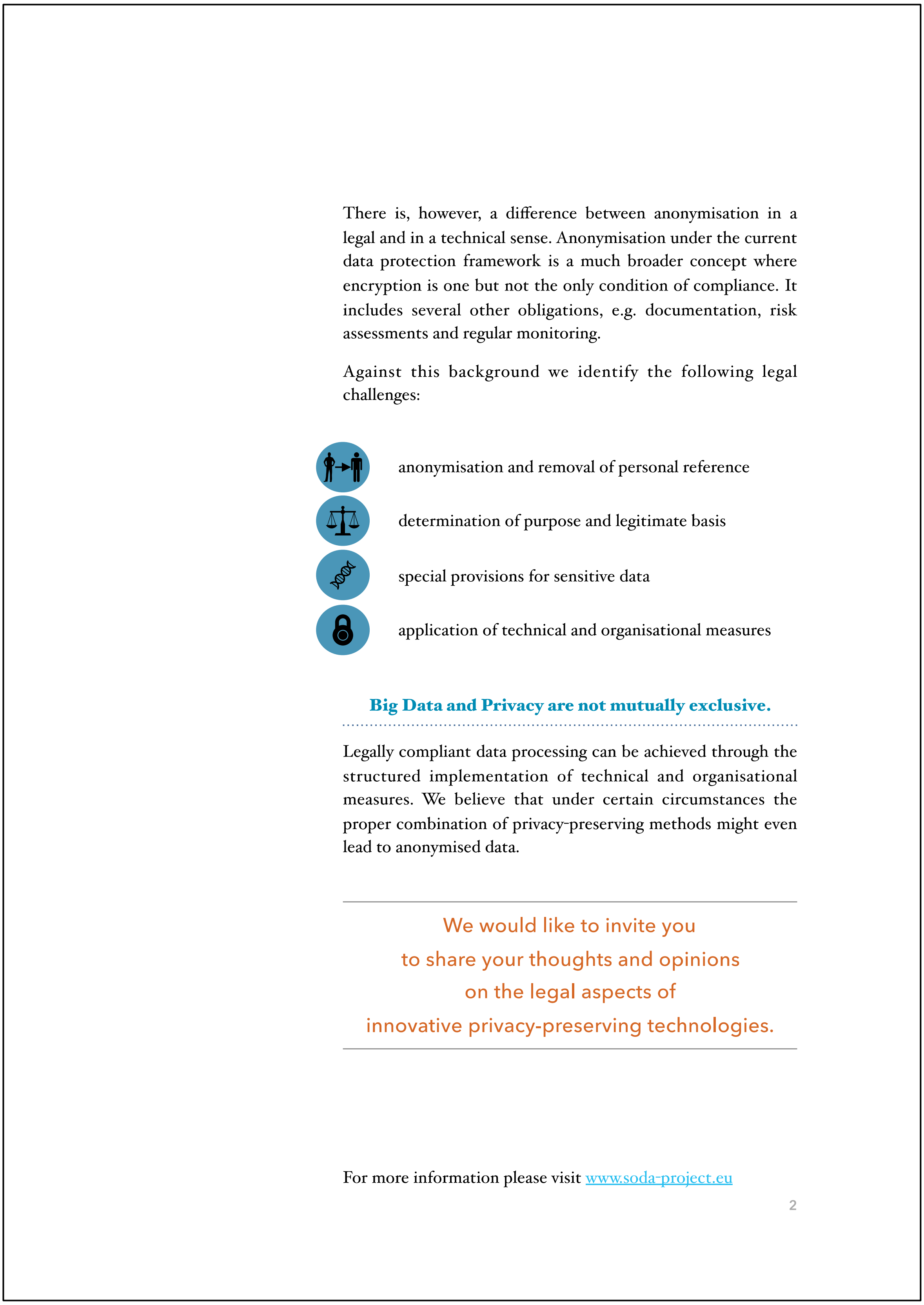
Our job as legal professionals is to work alongside researchers in computer science on the technical side and various other stakeholders, so that we can provide an accurate, in-depth legal analysis of international and European data protection and privacy laws, especially but not limited to the GDPR.

De-identification reduces risks and enhances privacy

Cryptographic solutions, such as multi-party computation and differential privacy have the potential to de-identify personal data in a way that does not allow the data subject to be identified with means reasonably likely to be used.

These state-of-the-art technical measures safeguard the privacy of the data subjects not only during but also after the data analysis.

1



There is, however, a difference between anonymisation in a legal and in a technical sense. Anonymisation under the current data protection framework is a much broader concept where encryption is one but not the only condition of compliance. It includes several other obligations, e.g. documentation, risk assessments and regular monitoring.

Against this background we identify the following legal challenges:

- anonymisation and removal of personal reference
- determination of purpose and legitimate basis
- special provisions for sensitive data
- application of technical and organisational measures

Big Data and Privacy are not mutually exclusive.

Legally compliant data processing can be achieved through the structured implementation of technical and organisational measures. We believe that under certain circumstances the proper combination of privacy-preserving methods might even lead to anonymised data.

We would like to invite you
to share your thoughts and opinions
on the legal aspects of
innovative privacy-preserving technologies.

For more information please visit www.soda-project.eu

2