MAX VON GRAFENSTEIN

Reconciling Conflicting Interests in Data through Data Governance

An Analytical Framework (and a Brief Discussion of the Data Governance Act Draft, the Data Act Draft, the AI Regulation Draft, as well as the GDPR)

HIIG DISCUSSION PAPER SERIES 2022-02

ABSTRACT

In the current European debate on how to tap the potential of data-driven innovation, data governance is seen to play a key role. However, if one tries to understand what the discussants actually mean by the term data governance, one quickly gets lost in a semantic labyrinth with abrupt dead ends: Either the concrete meaning remains unclear or when an explicit definition is given, it hardly describes the challenges, which are considered essential in this article, at least within the highly regulated EU Single Market. The terminological and conceptual ambiguity makes it difficult to adequately describe certain challenges for data governance and to compare corresponding solution mechanisms in terms of their conditions for success. This article, therefore, critically examines and further develops elements of data governance concepts currently discussed in Information Systems literature to better capture challenges for data governance with particular respect to data-driven innovation and conflicting interests, especially those protected by legal rights. To reach this aim, the article elaborates on a refined data governance framework that reflects practical experience and theoretical considerations particularly from the field of data protection and regulation of innovation. Against this background, the outlook briefly assesses the most relevant current draft laws of the EU Commission, namely: the Data Governance Act, the Data Act and the AI Regulation (especially the last one concerning the General Data Protection Regulation).

KEYWORDS

data governance, regulation, data-driven innovation, value and risk of data, data intermediaries, European data strategy, Digital Services Package, Data Governance Act, Data Act, AI Regulation, GDPR

CITATION

Grafenstein, M. v. (2022). Reconciling Conflicting Interests in Data through Data Governance: An Analytical Framework (and a Brief Discussion of the Data Governance Act Draft, the AI Regulation Draft, as well as the GDPR). HIIG Discussion Paper Series 2022-2. 45 pages.
https://doi.org/10.5281/zenodo.6457735.

AUTHOR INFO

RA Prof. Dr. Max von Grafenstein, LL.M.
Email: max.grafenstein@hiig.de

AFFILIATION

Alexander von Humboldt Institute for Internet and Society, Französische Straße 9, 10117 Berlin

Einstein Center Digital Future, Berlin University of the Arts, Berlin Career College, Postfach 120544, D-10595 Berlin

CONFLICT OF INTERESTS

The author certifies to have NO affiliations with or involvement in any organisation or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

ACKNOWLEDGEMENTS

Felix Biessmann, Daniel Fürstenau, Rebecca Frank, Ron Neumann, Christopher Olk, Sebastian Offermann, Jörg Pohle, Juliane Stiller, Rita Streblow, Violeta Trkulja, Lena Ulbricht, Alina Wernick, Anne-Katrin Witte, and Sören Wortmann

FOREWORD

This analysis is intended as a contribution to the data governance discussion, to analyse the suitability of different data governance models and other contributions and to help compare them in order to address the challenges associated with the topic. Since the discussion paper has the character of a preliminary research result, all readers are particularly invited to criticise the framework and, at best, to develop it further. The paper is, thus, intended as a living document, which the version history is intended to help in a documentationally transparent way.

Version	Date	Changes	Editor
1.0	06.04.2022	-	v. Grafenstein

CONTENTS

EXECUTIVE SUMMARY

Aim of data governance and its challenges

- Data governance aims, from a multi-stakeholder perspective, to reconcile conflicting interests in data (use), which diverge amongst the different stakeholders involved in terms of its value and risks.

- Perceived risks of using, especially sharing, data can be manifold, e.g. economic disadvantages or reputational damages, however, in highly regulated markets such as the EU, compliance risks also play a decisive role (e.g. violating data protection law or trade secret protection); "data quality" thus means that data is not only technically but also legally "fit for use".

- In contrast to those concrete risk perceptions, the value of data remains abstract and vague unless there is a specific use case, which is, however, often not yet determined by the time the data is accessed, especially in data-driven innovation.

- These different perceptions of risk and value pose a significant challenge to data sharing since data holders usually only share their data voluntarily when the risk of sharing the data is significantly lower than the value proposed by users interested in the data. To solve this value-for-risk dilemma, stakeholders need to balance the risks and value of the data (use) through appropriate data governance structures and processes so that data use is worthwhile for them.

- Another challenge arises from the fact that the value and risks constantly change over time depending on how the data is specifically used, which requires flexible data governance structures and processes constantly adapted over time.

- To cope with these challenges, the stakeholders involved in the use and re-use of data must coordinate on all three data governance layers, i.e. the regulatory, organisational and technological layer. This is an extremely complex task, not only because of the different goals, methods, processes and structures of the respective actors, but also because of their different mental models and terminologies.

- Last but not least, there is a conflict of goals that arises on all three data governance layers from the degree of participation or centralisation: here, inclusivity and multi-stakeholder control are set against an increasing complexity of structures and processes.

Recommendations to the EU legislator of the Digital Services Package

- With sharing obligations and access rights, the EU legislator solves the value-for-risk dilemma, but not the question of how the protection laws can be met, which must still be respected. To further strengthen successful data governance in practice, the EU legislator should therefore make the following adjustments in its draft laws:

    - In the Data Governance Act (Chapter II), the harmonisation framework for the publication of protected data held by public bodies should be extended to include an obligation to catalogue such data (and publish the catalogues) and a complementary right of access for everybody, provided that protection rights are respected; the competent authorities must not only be

authorised and obliged to systematically analyse the conflicts of interest that inevitably arise due to these access requests but also to proactively develop (and publish) solutions – this requires sufficient resources (in current data protection enforcement, the lack of concrete solutions due to the scarcity of resources is a major problem).

- The right of access to data between private parties in the Data Act (Chapter II) should be extended to a general data access right under a blocked period of five years; the block is to be automatically lifted after five years, so that the general access right comes into effect if the following conditions are met:
  - No structures or processes have yet emerged between private parties according to which the data sharing dilemma could be satisfactorily overcome on a voluntary basis (i.e. insufficient corresponding risk minimisation measures or value realisation mechanisms).
  - In the course of the five years, the competent authorities have succeeded in systematising conflicts of interest that typically arise with regard to usage data (which, according to the current draft law, are already subject to a corresponding sharing obligation) and in developing (and publishing) suitable solution measures.
  - In five years, it was also possible to clarify whether and how successfully an allegedly infringed party can prove a misuse of (usage) data with the consequence that the opposing party must destroy its products resulting from the proven misuse of such data.
- So-called gatekeepers should not be granted data access rights; the data sharing obligation for SMEs should be exempted. All costs of private data holders must be reasonable and should be borne by the data user, whereby SMEs (as well as non-profit organisations etc.) should be privileged to bear only the costs directly related to making the data available.

- Due to their central position and, thus, their accumulated knowledge and economies of scale, data intermediation services are particularly suited to assess the technical and organisational means to comply with protection laws. In this respect, they also form an important counterpart when dealing with the (possibly restrictive) solution practice of the competent authorities. Furthermore, they play a special role in overcoming the value-for-risk dilemma, insofar as there are no access rights, but access is based on the voluntary decision of the data holder. Regarding the special case of data sharing services, the legislator should align Chapter III of the Data Governance Act more stringently with the concrete needs of the actors involved and clarify its ratio:
  - First, a need for a verified trust signal exists for data holders and data users when they use a sharing service that they cannot fully control due to a lack of full instruction authority. A notification duty of sharing services should thus only apply to services that process the data for its own purposes (see or compare the role of "controller" and "processor" according to Art. 4 no. 7 and 8 GDPR).
  - Second, instead of generally prohibiting sharing services from using the data they transmit for their own purposes, they should only be obliged to a) make the risks caused by their own processing purposes transparent to the data holder and data user (insofar as this is not already required by the GDPR or the Digital Services Act) and b) subject themselves to a mandatory

certification procedure or code of conduct; c) only gatekeepers should be prohibited from using the data for their own purposes.

  - Third, the legislator should open up the possibility of registration under Art 15 Data Governance Act for not-for-profit data users as well, because, like non-profit sharing services, they have a need to communicate their non-profit orientation to the outside world through a verified trust signal.

- Last but not least, the AI Regulation should be better aligned with the GDPR and the Data Governance Act to avoid both over-regulation and a lack of regulation:

  - The AI Regulation's material scope as well as the individuals' fundamental rights-oriented risk-based approach are largely the same as the GDPR. Thus, their protection approaches and instruments (especially conformity assessments) can and should be more seamlessly aligned.

  - The focus on individual fundamental rights risks and thus the overlap of material scopes falls short given that there are more systemic risks that arise from AI systems; this applies at least to systemic risks in the field of nature and environmental protection (which, unlike most other areas of application listed in the AI Regulation draft Annex II, cannot be addressed by data protection laws through referring to "personal data"). The legislator should re-consider this.

  - However, the AI Regulation draft complements the GDPR well, at least in that the draft introduces the long-demanded liability of IT developers.

  - Another good aspect of the AI Regulation draft is the required exchange of information with regard to new risks that may arise. However, this regulatory mechanism should not only apply to users of AI systems, but to all kinds of data users (since the fundamental problem of unforeseeable risks arises for all kinds of later data uses). Thus, data users should generally be obliged to constantly record factors of such new risks in the data catalogues and systematically collect knowledge about it and make it available to up and downstream users (as well as data holders, given that shared responsibilities require this). However, the right place for this obligation would be the Data Governance Act, as it is a general requirement to maintain data quality and does not only refer to risks of an AI system or of processing personal data.

Recommendations to further stakeholders

- Private stakeholders can actually use the mutual dependencies between the regulatory, organisational and technological layer as a competitive advantage. This is especially true when their buyers, contractors, etc. are held legally (or at least politically) accountable to meet certain standards, but technically only their vendors or providers are able to do so. These sellers or providers can then take advantage of this circumstance and offer such products or services that can be used according to those standards.

- The place where these coordination efforts accumulate is the data catalogues, where the actors involved must add their constantly generated knowledge about the value and risks of the data and the necessary technical-organisational measures. Therefore, stakeholders, should also focus on the creation of such shared and flexibly adapted data catalogues.

1 INTRODUCTION: HOW TO IDENTIFY SUCCESSFUL DATA GOVERNANCE (ESP. IN THE EU SINGLE MARKET)?

In the current European debate on how to tap the potential of data-driven innovation, data governance is seen to play a key role. The increasing amount of data is considered as an important factor for better decision-making in the public and private sector. Since data can be reproduced almost free of charge, in principle, data can also be used by an unlimited number of natural or legal persons without disadvantage for the other. Not only the collection but also the sharing and re-use of data, therefore, promises to contribute to an important part of the value creation for individuals, companies and public welfare. Reality, however, seems to fall short of this claim, data remains in their silos and is shared much less often than desired.

1.1 How the EU Digital Services Package seeks to help

The European Commission sees one reason for this in weak data governance structures, and consequently stresses the necessity for better data governance concepts in Europe for certain sectors (so-called data spaces) as well as cross-sectorally (European Commission 2020a, pp. 1–13). Especially with its draft for a Data Governance Act from the 25th November, 2020, the Commission seeks to set up "governance structures and mechanisms that will create a coordinated approach to using data across sectors and Member States" (European Commission 2020b, p. 2). However, despite the importance the Commission sees in better data governance, neither the European data strategy nor the draft of the Data Governance Act define the exact meaning of this term. Reading the legal text, at least, it becomes clear that data governance mechanisms can be of legal, technical or organisational nature.[1] Similarly, in its second major initiative, i.e. the Data Act proposal from 23rd February 2022, the Commission aims at "ensuring fairness in the allocation of value from data" by fostering access to and use of data (European Commission 2022, p. 2) while, at the same time, respecting conflicting rights by requiring the actors involved to implement "all reasonable technical, legal and organisational measures" (European Commission 2022, p. 16). Such conceptual ambiguity not only causes theoretical but also practical problems: As long as it remains unclear how the aforementioned legal, technical and organisational aspects interrelate, it is difficult for the EU legislator, as well as for any other entity involved in the collection, sharing or re-use of data, to better assess and verify the impact of their own contributions to "better data governance" and, thus, to their own interests, those of other persons or the public in data-driven innovation.

1.2 Main challenge for successful data governance

The last consideration points to a major conflict of objectives, on which this paper focuses. As a non-rival good, as mentioned before, data can be used by an unlimited number of persons, without disadvantage for the other, however, this only applies in principle. In reality, the collection, sharing and re-use of data can conflict with so many interests (Gangadharan 2014), which in the EU are often also legally protected, meaning that the risk of violating a law or suffering some other disadvantage seems, in the eyes of many stakeholders, to outweigh the expected value (Friederici et al. 2019, p. 38). The result in day-to-day data re-using and sharing practices fall short. More specific reasons for this may be general risk aversion of the stakeholders involved, their fear of reputational damage, and/or because they cannot afford larger legal departments to address the legal issues (see for the latter Frey & Presidente 2022; von Grafenstein 2020a). To address this challenge, a concept of data governance must clarify how data governance particularly interrelates with both the enablement of data-driven innovation and full respect of conflicting interests, especially when they are protected through legal rights. Only on such a conceptual basis is it possible to

identify typical conflicts of interest for individual sectors or across sectors and thus to define the corresponding governance mechanisms as so-called data spaces or even cross-sectoral.

1.3 Current state of research

This is, indeed, where many research approaches step in: In IS literature, a commonly seen purpose of data governance is to increase the value of data and minimise data-related cost and risk. To reconcile these two conflicting goals (i.e. exploiting value and minimising cost and risk), data governance especially specifies and formalises decision rights, procedures and controls (Abraham et al. 2019, pp. 424-426). This definition is able to describe the conflict between an expected value of collecting, sharing and re-using data and a corresponding risk, e.g. for a legally protected interest or, vice versa, a compliance risk. However, when analysing the relevant publications, what remains unclear is the specific challenges faced in order to reach this goal. Most publications focus on certain aspects that this contribution also considers to be relevant, however, there is no framework that represents the particular interdependencies of the legal, organisational and technological aspects together. Abraham et al, for instance, develop on the basis of a structured literature review of 145 research and practitioner publications (published during 2001 and 2019) on a data governance framework, which essentially focuses on the organisational aspects of data governance (Abraham et al. 2019, pp 425 et seq.). However, this framework does not make clear the particular challenges of these organisational data governance aspects mediating between data-driven innovation and legal regulations. One concept that makes this observation obvious is the term "data quality". In a nutshell, "data quality" means "fitness for use" (Otto 2011c referring to Wang 1998). However, many authors seem to assume a static and rather technical concept of data quality, which is independent of the respective context of use and of whether the data may also be used legally in this context. Krotova and Spiekermann propose, for example, a five-step-scheme assessing the value of data according the following order: 1) the relevance of data (especially whether certain laws apply), 2) the quality of data, 3) the costs of maintaining the data, 3) the usefulness of the data, 4) the data market value (Krotova and Spiekermann 2020). The authors obviously assume that the applicability of laws decides in advance of all further steps whether the data may be processed at all, while the data quality seems to be independent of the usefulness of the data. In contrast, the question is more about how the data may be used in specific contexts so that their value can be exploited.

The clearer a framework conceptually captures and addresses these challenges, the better the respective stakeholders will be able to define and establish the appropriate decision rights, procedures and controls. To reach this aim, this article ties in conceptual elements from data governance discussed especially in information systems by drawing on practical experience and theoretical considerations, mainly from the field of data protection as well as the broader area of regulating innovation. Each chapter will focus on another aspect by refining the terms and concepts used in the current discussion as described. In doing so, the paper does not claim to provide a "full theory", i.e. a comprehensive and detailed explanation of the "data governance" phenomenon (cf. Hassan et al. 2022). Rather, by clarifying and developing individual but interrelated key terms and concepts, this paper aims to provide a theoretically more solid basis for describing certain challenges and conflicts in the field of data governance more precisely than before and, thus, for examining and comparing the appropriateness of potential solutions. In conclusion, the framework aims to provide a common basis for interdisciplinary research in the field of data governance and, therefore, more effective public and private regulation that is evidence-based. Against this background, the outlook briefly analyses the most relevant current draft laws of the EU Commission in terms of how well they help to meet the data governance challenges described, namely: the Data Governance Act, the Data Act and the AI Regulation (especially the last one with regard to the General Data Protection Regulation).

2 ANALYTICAL DATA GOVERNANCE FRAMEWORK

2.1 Data *governance*: Reconciling conflicts of interest in data through coordination

A first refinement concerns the clarification of the term "governance". As said before, data governance is usually defined as the exercise of authority and control over the management of data through specifying and formalising decision rights, procedures and controls (Abraham et al. 2019, pp. 424–426). This understanding refers, in more abstract terms, to situations in which an entity aims at intentionally steering events and behaviours through certain formalised measures. Instead, in governance and regulation research, the aforementioned definition corresponds to the understanding that many researchers have of the term "regulation", as a sub-category of "governance". The term "regulation" in this context refers to a public or private entity (i.e. the "regulator") that aims at causing or maintaining a certain situation or behaviour through certain formal or informal measures (Black 2014; with respect to algorithmic regulation, see Yeung 2017). In contrast, the term "governance" usually has a broader meaning (see the overview of possible definitions, for instance, at Braithwaite et al. 2007, p. 3). According to the political scientists Hofmann, Katzenbach and Gollatz, for example, internet governance focuses less on mechanisms of control of one entity over others, but rather on reflexive coordination amongst the entities (Hofmann et al. 2017).

A reason for the narrower understanding of (data) governance in the IS discourse may be because most data governance reflections still focus on intra-organisational data governance; there is little knowledge about how data may be shared in inter-organisational relationships, especially given the problems raised by data protection law (Abraham et al. 2019, p. 433; see more recently, Janssen et al. 2020, p. 6). Thus, it is not surprising that this discourse has chosen a governance definition that primarily reflects the perspective of one entity asking how this stakeholder might exercise authority and control over its own data. However, to make data sharing work, especially, on a society-wide level, or at least in certain sectors, it is more a question of the coordination between different stakeholders than a question of authority and control of one of them. This coordination of several actors to reconcile their different interests is one of the central challenges for successful data governance, as the following chapters of this paper will repeatedly demonstrate. In sum, the definition proposed here is only a slight broadening of the original perspective. However, this broader understanding of data governance ensures that the analysis is not narrowed prematurely to a one-dimensional control approach, and, therefore, the decisions to be made do not miss the actual problem. Researching data governance by focusing on coordination and its multiple perspectives, thus, aims at providing a more comprehensive knowledge base for public and private entities to make better regulatory decisions.

2.2 Roles of actors involved: Data holders, data users, etc.

When analysing how different stakeholders coordinate to reconcile their interests in the collection, sharing and re-use of data, it is necessary to clarify their roles. The following figure may give a first overview. The two primary actors are "data holders" and "data users" (cf. European Commission 2020a, p. 7). The term "data holder" refers to an entity (i.e. a natural or legal person as well as a specific department or employee) that has legal control or, where the law does not assign control to any entity, de facto control over the data (cf. Art. 2 no. 6 Data Act Proposal). In contrast, "data users" are those entities who use or reuse data, or more specifically, are interested in the use of data that the other entity may hold. The term "data user" used here with its reference to an interest in the data, therefore, differs from the terms "data users" and "data recipient" according to Art. 2 no. 6 Data Governance Act Proposal and Art. 2 no. 7 Data Act Proposal. The reason for this is that both draft laws apparently declare the relationship of interests to be "now clarified"

and, therefore, no longer have every interested entity in mind, but only the entities that are now legally authorised. The focus used here, to the contrary, declares the relationship of interests to be still unresolved. This is necessary to keep open the question of whether legal measures like the draft laws can adequately clarify the conflicts of interests in the data. Of course, entities can also play a dual role as data holders and users of data (see already Wernick et al. 2020).

If data holders allow data users to access their data, the coordination efforts of both entities may be orchestrated by an inter or intraorganisational intermediary. Such an intermediary could take any form, ranging from a purely virtual form to an independent legal entity. Further stakeholders involved in (or concerned by) the data treatment may also participate in (or contribute to) such an intermediary role. Thus, the term "intermediary" used here is again broader in meaning than the term used in the Data Governance Act proposal, which mainly focuses on intermediaries "that are independent from both data holders and data users" and "assist both parties in a transaction of data assets between the two". The subsequent sections will describe several further roles involved at different layers of data governance.

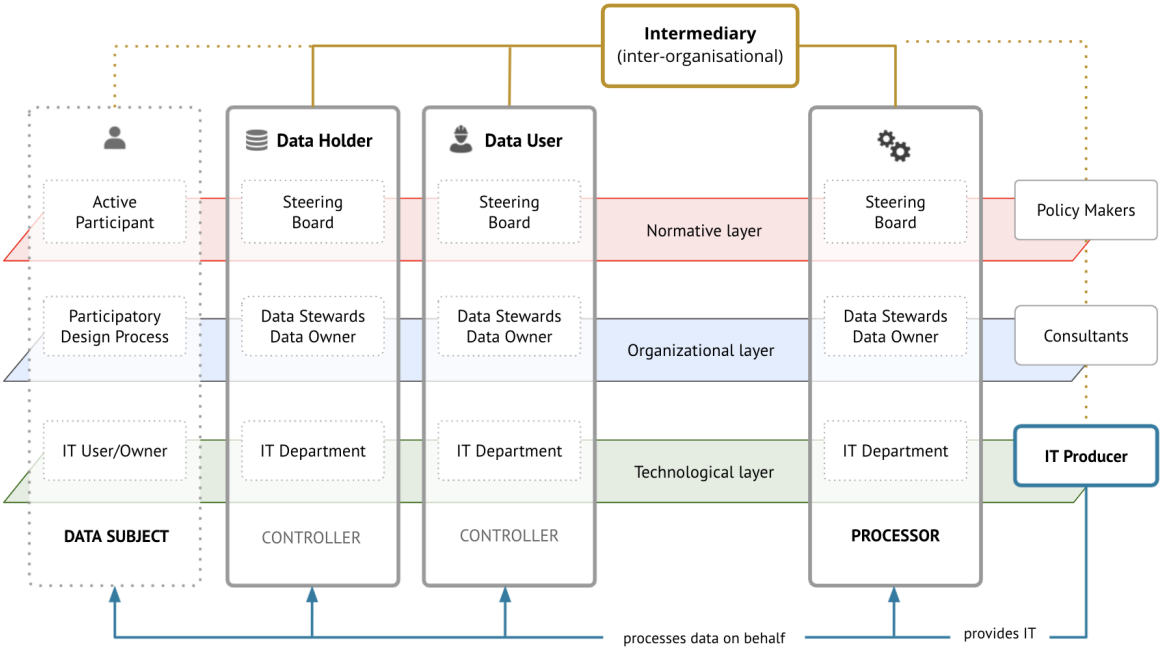

Fig. 1: Overview of data governance layers and actors.

2.2.1 Note on the terms taken from data protection law (processors and data subjects)

A few aspects should be clarified with respect to data protection law: Albeit different laws apply to non-personal data and personal data, this contribution refers to the term "data" implying both subcategories. The term "personal data" is used only to highlight an aspect relating to data protection law. In this case, the notion "data subject" is used to describe the natural person to whom the data relates (Art. 1 and 4 sect. 1 GDPR). If data subjects play an active role in the governance of "their" data (i.e. either withholding or using the data), they are also referred to as data holders or data users. Furthermore, the relevance of personal data also plays a role with respect to data protection rights (e.g. the right to data

portability, Art. 20 GDPR) or duties (such as the requirement to retrieve the consent from data subjects as a legal basis for the processing, Art. 6 sect. 1 lit. a GDPR). In contrast, even where data protection laws are applicable, this framework does not use the legal-technical term "controller" but usually refers to its subcategories, data holders and/or data users, because they are more precise in the context of data governance. In data protection law, a "controller" defines the purpose of data processing and its principal means, Art. 4 no. 7 GDPR (and is thus the entity that is primarily responsible for legal compliance). However, the terms "data holder" and "data user" both imply that these entities process the data for their own purposes. In order to clarify a situation where an entity gets into contact with data and does not process the data for its own purposes, but rather on behalf of another entity (i.e. a data holder or data user), the term "processor" is used (which is another common term in data protection law, Art. 4 no. 8 GDPR). An example is a cloud service provider that stores data on behalf of a data holder or if a software-as-a-service provider analyses data for a data user. Of course, the moment a processor processes data for its own purpose, that entity itself becomes a data holder or data user.

2.2.2 Further actors within and between data-processing entities (data owners, data stewards and steering boards)

Regardless of an entity's role, (i.e. as data holder, data user, or processor), the proposed framework refers to the common ontology in IS data governance literature that differentiates, mostly focusing on an entity's internal data governance, between "data owners", "data stewards", as well as data governance "steering boards" (Abraham et al. 2019, pp. 428–429). While data stewards (also called "custodians") are responsible for the data management, often in specific departments or areas, and assist other persons in the treatment of data, data owners determine the requirements for the data in their department or area, i.e. what qualities the data (must) have in this area. Thus, while data owners usually represent the department or area within an entity (e.g. as "head of"), stewards are formally responsible for the actual data treatment (Otto 2011a, p. 236 referring to Loshin 2008 and Khatri & Brown 2010). However, since the data owners are usually fully occupied with the management of the entire department, data stewards not only technically manage the data treatment but also have an important coordination function. Data stewards usually coordinate with stewards from other areas to implement and maintain common data governance principles across the entity as well as between several different entities. In addition, data stewards coordinate the diverging interests in the data across different departments and divisions. These coordination functions can become quite complex depending on the amount and variety of interests that they have to coordinate (e.g. from marketing and sales, research and development, human resources, legal, finance, IT, etc.) (Ladley 2019, p. 30). The resolution of conflicts, directing and monitoring, also make it necessary to create a certain hierarchy (Ladley 2019). In most cases, there is at least a data governance steering board (or "committee") that includes representatives from all departments and divisions as well as from various levels that are necessary for (or concerned by) the data governance activities (Smallwood 2014, p. 35). Smallwood sums up the objective of such cross-functional boards as: "The result is not only more secure information but also better information to base decisions on and closer adherence to regulatory and legal demands" (Smallwood 2014, p. 27). As mentioned previously, such an intermediary function can also be performed through appropriate mechanisms for data exchange between different entities acting in a common sector and even cross-sectorally (Wernick et al. 2020).

2.2.3 Actors contributing to the treatment of data without direct contact to data (e.g. IT developers)

Last but not least, there are entities that do not come into direct contact with data at all, rather they only provide data holders, data users, processors, as well as data subjects with the products, services and

frameworks that the latter need for their processing activities. For instance, there is the legislator, who contributes laws concerning the collection, sharing and re-use of data on the regulatory layer, but also further "policy makers" (in the broadest sense), such as enforcement authorities, legal courts, research institutions, think tanks, lobby groups, and so on. On the organisational layer, one may observe external firms providing consultancy on how to set up organisational structures and processes to exploit the value of the data, mitigate the risks, or make the normative and technological layers fit each other. Finally, on the technological layer, the proposed research framework calls such entities, which do not come into direct contact with the data but nevertheless contribute to its governance as "IT developers" that provide the technologies for the processing. In this context, it should be pointed out (in advance) that especially IT developers, but also other entities, may leverage the interdependencies that arise between the stakeholders involved at different levels as a business opportunity. The reason for this is that their customers (i.e. data holders, data users, processors, as well as data subjects) need technology and/or support regarding the regulatory and the organisational layer that enables them to reconcile the conflicting interests in the (reuse of their) data (von Grafenstein 2020a). The following chapters will focus on repeatedly illustrating these dependencies.

2.3 Coordination on a regulatory, organisational and technological layer

The previous chapter has already referred at various points to different data governance layers, namely the regulatory layer, the organisational layer and the technological layer. The legislator also refers to these three layers but without further explanation.[1] This chapter explains why these three analytical layers are suitable for describing the challenges of successful data governance, which arise specifically from the interdependencies of the actors involved (von Grafenstein et al. 2019, pp. 231/232). The chapter begins with a brief summary of the impact of the regulatory framework on resolving the conflicts of interest in data use.

2.3.1 Impact of regulation on data use (and governance concepts)

All the aforementioned stakeholders may perceive the value and risk of the collection, sharing and/or re-use of data differently. These differences may depend on their respective role, context-specific knowledge or perception. Thus, the value and risk of data is not objectively given but depends on the perspectives of the actors involved. One part of data governance, therefore, means to decide on which perspectives to consider, in particular, which means are used for considering these perspectives and for reconciling the corresponding interests (Günther et al. 2017, p. 197; Abraham et al. 2019, pp 430–431). To synchronise these expectations, in particular, with respect to the risks, and the measures that are necessary to control these risks, the law plays an important role. This especially applies to the EU Single Market, which has a high density of regulation compared to others (Matthijs et al. 2021).[2] The law is repeatedly used here as a trust-building framework that makes the free exchange of goods and services possible in the first place (cf. European Commission 2020c, p. 29). However, the trust-building effect of the law does not only exist in legally protecting certain interests per se, but also in synchronising the respective expectations. Protecting certain interests through laws makes these interests more objectively perceivable by other parties.[3]

Despite the importance of the law for the collection, sharing and re-use of data, few data governance concepts go into the details of how data governance structures should best reflect the legal framework. This is astonishing, given that laws like Basel II and the Sarbanes-Oxley Act have been key drivers for information governance and data governance, both in practice and conceptual discussions (Khatri & Brown 2010, p. 148; Tallon et al. 2013, p. 159). However, while both laws required banking and credit industry

players to collect, process and maintain data for specific purposes (especially for loan loss risk assessment), today's laws often have the opposite effect - not collecting, not sharing and not re-using data, at least, not for new purposes. The most common examples of this are privacy concerns (Günther et al. 2017, p 198; Foster et al. 2018, p. 1418; Janssen et al. 2020, p. 4) but also copyright (Foster et al. 2018, p. 1418) and fear of competitive disadvantages (Günther et al. 2017, p. 198). Even where laws explicitly require the sharing of data (Janssen et al. 2020, p. 4), implementation is, in practice, hesitant, and scientific approaches with constructive proposals are rare or, as in information science, focus on questions of what risks people in companies or public authorities see so that they do not share data (Frank et al. 2022). Despite the explicit goal of most data governance approaches to maximise the value of data and limit associated risks, most approaches do not go beyond mentioning the influence of the legal framework on the use of data (Abraham et al. 2019, p. 432, as well as Krotova and Spiekermann 2020; however, also see the positive examples of organisational measures listed by Janssen et al. 2020, p. 6). This is true even for publications that explicitly address the question of how data governance can create trust in the outcomes of data processing (Brous & Janssen 2020). Some authors, after all, see the usability of data as dependent on the rights, responsibilities and obligations associated with the information it contains (Beynon-Davies and Wang 2019, pp. 487 et seq.). So far, the observation of the EU Commission seems to be correct in that current data governance concepts are obviously not sufficient to facilitate data sharing. Therefore, to find the right data governance concepts, one must understand in greater depth how legal regulation interrelates with the processes and structures of data processing entities.

While legislation (e.g. Basel II) was originally a key driver for data governance, today laws create significant compliance risk in data sharing.

2.3.2 Examples from data protection law, and more

To demonstrate how neatly the law interrelates with an entity's organisational structures, processes and its technological design, data protection law gives an illustrative example. A starting point is the so-called "data protection and security by design"-approach under Art. 25 and 32 GDPR. The approach requires entities that process personal data (i.e. data controllers and their processors) to implement the legal requirements into the technological and organisational design of the data processing. Typical examples for such technological and organisational measures are anonymisation and encryption technologies (on the technological level), role-based access control and pseudonymisation (on the organisational level), non-disclosure agreements and similar data-sharing contracts, work instructions (on the legal-organisational level), and so on (Hansen 2019b Rn. 33-35; Hansen 2019a; Petri 2019; Bygrave 2020; Docksey 2020). What makes the data protection by design approach difficult to implement in practice is, as observations from the field suggest, that the approach requires considerable coordination efforts on the part of the stakeholders involved. A general reason for this is that the approach forces the regulation addressees to align their legal, organisational and technological systems with each other. Due to the systems' own logics and languages, this requires considerable translation efforts, which in practice are inevitably associated with corresponding translation problems, losses and costs (Hölzel 2019). Another more specific reason is that, in practice, legal responsibility and technical capabilities usually diverge. While the main legal responsibility lies with the controller, (i.e. the entity who determines the purpose of the processing), the controller usually relies on third-party providers that either process the data on its behalf (as processors, for example, providing "software as a service")[4] or provides the IT without coming into contact with the data (i.e. as so-called IT producers) (Vollmer 2021, sent. 4). In both cases, there are different entities involved who, on the one hand,

The governance challenge is that those who are legally (or politically) responsible for risks and those who can technically control those risks are often different entities or people.

bear the main legal responsibility, and on the other hand, are technically capable of applying the legal requirements. This means that these different entities must cooperate on different layers (at least on the legal and technological layer) in their different roles to ensure that a controller finally fulfils all regulatory requirements as the primarily responsible entity among them. From the perspective of the respective actors, these dependencies suggest different contributions regarding the necessary cooperation: Data controllers may, for example, check whether services or technological components from third parties, on which their data processing builds, enable them to meet the regulatory requirements. Vice versa, the better third parties enable a controller to meet the regulatory requirements, the better they may use this as a business opportunity and generate a competitive advantage (von Grafenstein 2020). In this respect, the GDPR can indeed be seen as having the potential to promote innovation, rather than as a barrier to innovation (von Grafenstein 2022). However, this requires a corresponding knowledge and mindset.

Even if there are already some positive examples in the meantime, as far as practical observations go, the required knowledge and mindset is not yet widespread in current practice. For example, the supplier may not give sufficient thought to how to enable buyers to comply with the regulatory framework applicable to them. In such a case, legal compliance is supposed to remain the other party's problem, especially because one does not want to get close to corresponding liability risks. In principle, this problem could be solved by means of appropriate agreements between the entities or by involving a third party that takes over the legal or at least financial liability (e.g. a certification company or an insurance company as is becoming increasingly common in the case of cyber security incidents). However, these mechanisms have not yet become established in the market. Apart from this, coordination also often fails because the various departments within a company would have to coordinate their efforts. For example, the marketing and sales department would have to use legal arguments in their sales strategy and reconcile these arguments with the legal views from the legal department (which salespeople are not always comfortable with). To do this, the legal department would have to work with the development department to ensure that the products sold actually contribute to legal compliance on the part of the customer. For this to work, lawyers would have to abandon a reactive approach auditing a product after development has been completed and apply, instead, a more proactive approach actively contributing to the development of products with concrete proposals for solutions (which lawyers are not trained for, at least not yet on a large scale). The management would finally have to decide that making customers legally compliant is part of the company's business strategy (which is also not yet very common) and allocate resources for the necessary awareness, learning and coordination efforts. In fact, most companies are still in the early stages of such strategic alignment. The fact that the EU Commission sees itself as a pacemaker in this respect confirms this observation (European Commission 2020a, pp. 17 et seq.).

Actors from different disciplines involved in data sharing need to align their terminology, metrics, methods, processes, and responsibilities.

The preceding example illustrates how neatly data protection law interrelates with the organisational and technological layer of data processing entities. However, the same dependencies exist with respect to other legally protected interests, such as trade secrets and intellectual property, and even further social, economic, political or public interests. According to Art. 2 sect. 1 lit. c Trade Secret Directive, the legal protection of trade secrets requires, for example, "the person lawfully in control of the information" to take "reasonable steps under the circumstances (...) to keep it secret". Whilst on an organisational layer, a data holder must categorise and document its protected information and sensitise its employees. An IT provider may enable the data holder to encrypt the information on a technological layer; furthermore, if the data holder passes on the information to another data user, the data holder must legally oblige the user of that information to equally bind its employees by a non-disclosure agreement (NDA) (Sarkar et al. 2018, p. 226). Even in a

purely political context, a municipality may only meet the expectations of its citizens, for example, to improve the air quality, if its IT provider takes the regulatory and organisational antecedents of the municipality into account.[5] Given citizens' rights of access to environmental information (2003, Freedom of access to information Directive), for instance, the provider may enable the municipality to fulfil these access rights. The provider may also help to organise the local-political debates that are likely to ensue around the appropriate measures to reduce the pollution. Of course, the provider is usually not legally obliged to do so. However, if the provider wants to increase its sales opportunities, or solve the societal-environmental problem, it should not only sell its "technological solution" but also enable its purchasers to holistically solve the problem with respect to the entire conflict of interest. Those examples may illustrate how different actors should coordinate on a regulatory, organisational and technological layer to make the collection, sharing and re-use of data work, (i.e. to reconcile their interests in the data). All this means, in conclusion, that the claim "making data fit for use" means to implement the necessary organisational and technological measures so that data users are also legally, or in a broader meaning, politically allowed to use the data for a specific use case. The term "data quality" therefore has a clear regulatory dimension.

2.3.3 Description of analytical layers

Against such a background, this paper prefers to distinguish three analytical data governance layers: the regulatory layer, the organisational layer, and the technological layer. It is important to emphasise that these layers are used in a purely analytical sense (therefore, one can also speak of dimensions): They are not meant to be exclusive nor the only correct ones. The proposed layers contain many elements that are mentioned in other analytical layers discussed in literature, which also partly overlap. In practice, all three layers are also closely interlinked. Therefore, the proposed layers are rather meant to complement existing frameworks (especially the data governance framework proposed by Abraham et al. 2019, pp. 424-438) by shifting the focus of attention to the specific challenges that arise when stakeholders seek to coordinate their interests by taking these regulatory, technological and organisational aspects into account. Against this backdrop, the proposed layers are described as follows:

The regulatory layer

The regulatory layer consists of the applicable law concerning the collection, sharing or re/use of data (cf. the macro-level at Foster et al. 2018, pp. 1418 et seq. and the external antecedents as well as the cultural and strategic factors at Abraham et al. 2019, p. 432). Often, it is unclear how the different laws relate to each other in a certain context. Therefore, the actors involved must first bring the regulations into a consistent framework with respect to their specific situation. Besides the applicable law, the regulatory layer also covers private ordering, e.g. contractual agreements, as long as these means are enforceable on the grounds of private law and therefore have (at least partly externally) a binding effect. Incidentally, besides legal means, cultural values and social norms can have a similar regulatory function (Abraham et al. 2019, p. 432; Janssen et al. 2020, p. 4). This is the case as long as an individual actor also perceives them as externally imposed and as a more or less socially sanctioned obligation.

Furthermore, the regulatory layer does not only look at these public, cultural, social and private norms per se but includes events and behaviours that are directly covered by these norms. This is necessary to describe how certain entities or persons transpose these norms into practice. Even if most regulations, for example, are set up under the assumption to be applied by the regulation addressees without any ifs or buts, these addressees are, in fact, free to act accordingly. This factual room for manoeuvre is an essential difference to regulation by technology, which is de facto more restrictive (Lessig 2006, p. 125). Of course, such rooms for manoeuvre vary from sector to sector. In the public sector, the regulation addressees, (i.e. public

authorities), are strictly bound by the law on the basis of which they may act, especially when intervening in the private sector; plus, they cannot invoke their own fundamental rights. As a result, the scope of action of the persons acting on behalf of an authority is much more narrowly defined by law than in the private sector. In contrast, private parties can invoke their fundamental rights, which structurally leaves them far greater room for manoeuvre (Papier 2006, § 55). Regardless of the specific regulatory strategy implemented by the legislator (e.g. command and control, co-regulation or self-regulation) (Baldwin et al. 2013), private parties may therefore follow different strategies on how to apply the law. From their perspective, the law is one factor among many that hinders or supports them in meeting their interests. Depending on their specific considerations in their situation, for example, private actors may disregard the legal requirements and hope for an enforcement deficit (let's call it the "reckless-approach"); or apply them as far as absolutely necessary and improve them if required (the "classic compliance-strategy"); or strictly apply them out of a sense of duty (whether as an act of "anticipatory obedience" or for fear of loss of reputation); or use them as a business opportunity and competitive advantage (which can be seen as "making a virtue out of necessity"), etc. (Günther et al. 2017; von Grafenstein 2020a).

Data strategy

These margins of manoeuvre are important because they enable the respective stakeholder to make what they perceive to be the optimal decision, or rather, the best compromise between certain conflicting interests. In current practice, few stakeholders already have a clear data strategy as to which approach is best suited for them to achieve their goals in their specific situations (i.e. to optimally resolve the conflict of objectives between maximum value and minimum risks). The strategies of the EU Commission and of the German government, which are based on extensive consultation processes, reflect this current state (European Commission 2020a, pp. 7-13; Die Bundesregierung 2021, especially on p. 7). In my opinion, the approach of understanding the needs of the stakeholders involved with respect to the applicable regulatory framework and using them as business opportunities and competitive advantage has great potential (although legislators would have to more consistently draft their laws in such a way).[6] In any case, only when the strategic approach is clear, one could design mechanisms in a way that is consistent with the entity's objectives. Therefore, the proposed framework analytically locates these strategic decisions on the regulatory layer, even if these decisions may flow seamlessly into their organisational implementation.

Turning to the organisational layer, this layer consists of all those structures, processes and practices that concretely implement the entity's data strategy on how to maximise the value of data and minimise the risks, in particular compliance risks. Abraham et al. consider these mechanisms as the core of data governance (Abraham et al. 2019, p. 427). Authors often distinguish between structural, procedural and relational mechanisms. Structural mechanisms build on the allocation of decision-making authority by assigning roles and responsibilities. Procedural data governance mechanisms encompass policies, standards, processes, performance measurement, compliance monitoring and issue management. Often, authors also categorise contractual agreements under these procedural mechanisms (Abraham et al. 2019, p. 427 with further references). However, as mentioned, the proposed framework in this contribution classifies these mechanisms at the regulatory level. Apart from that, authors finally list relational mechanisms that enhance the collaboration amongst stakeholders. Communication and training are example areas of where to raise awareness within, and eventually, between entities for the data governance program (Abraham et al. 2019, p. 427 with further references). A very first necessary step for setting up data governance structures and procedures are so-called data catalogues. In short, data catalogues are organised inventories of data containing metadata to help entities collect, organise, find, access and use the data.[7]

The organisational layer

The technological layer

Last but not least, the technological layer is defined by its architectural design consisting of the software and hardware infrastructure for processing the data. Often in data governance literature, the technological layer is not explicitly mentioned (see, for example, Abraham et al. 2019; in contrast, see Janssen et al. 2020, pp. 4–5, as well as Krotova & Eppelsheimer 2019, p. 8). This may be due to the attempt to distinguish the more recent data governance approach from the pre-existing IT governance discussion (Khatri & Brown 2010, p. 149; Tallon et al. 2013, pp. 142 et seq.) However, to adequately address the challenges for successful data governance that arise from the interdependencies and associated coordination efforts of the actors involved, the technological layer should be part of any data governance framework. The interplay of the three data governance layers becomes apparent not least when it comes to the question of how the stakeholders may specifically cooperate on all three data governance layers.

2.4 Dynamic value and risk of data: the data sharing dilemma

The stakeholders involved in the collection, sharing and re/use of data refer to data governance mechanisms that maximise the value and minimise the risks and costs that they respectively expect. However, the dynamic nature of the value and risk of data poses significant problems in finding the right point where the expected value exceeds the perceived risks and costs. This is especially true for data-driven innovation, where the value and risks only materialise in the course of the innovation process.[8] This dynamic represents a major challenge, in particular for data sharing, assuming that the stakeholders cooperate in the collection, sharing or re/use of data only if they expect the value to be higher than the expected risks and costs. To understand this challenge in more detail it is helpful to first distinguish between abstract and specific types of the value and risk of data. On this basis, it is easier to understand the problem that especially arises when data is transferred from one entity to another. Finally, against this backdrop, it is possible to determine whether exploiting the value of data and controlling its risks is worth the costs.

2.4.1 Interplay between abstract and specific value and risks

In the current data governance discussion, authors push the need for metrics, especially for a quantitative assessment of the value of data (Abraham et al. 2019, p. 434). The hope behind this claim often seems to be that if there is a price tag attached, data holders would be more likely to contribute to data sharing or that data subjects would be more justly compensated for sharing their data. In this case, the slogan is usually "data for money" (van de Ven et al. 2021). As justified as the demand for a quantifiable value of data may be, it should not distract from the fact that many data holders, as well as data subjects, have long attached a value to their data. In many cases, these values are indeed quantifiable, even if the stakeholders involved have so far attributed these values to the data indirectly through individual exchange relationships rather than via an objective market price. Therefore, the following remarks concentrate on determining what type of value and risk is actually attached to data in what situation. This may broaden the perspective on data governance solutions that, in turn, may make data sharing work.

To start with, most data is collected in a specific context for at least a technical purpose, which can actually be used to determine the value of the data. For example, private companies may set cookies in the browsers of their website visitors to show them personalised advertising on their sites. Even if this purpose does not directly reflect the quantitative value of the data (collected via the cookie), such a value can be concluded by the benefits expected from the stakeholders involved. For example, the visitors of the websites may, *nolens volens*, tolerate the cookies because this enables them to access the websites for free. The private companies, in turn, may set the cookies for personalised advertising to increase their profits. Thus, the website visitors

assess the value of the data with respect to the service they get in exchange for disclosing their data (von Grafenstein 2021c), while the website owners may assess the value of the data with respect to their business model (cf. Sorescu 2017). In both cases, the value of the data can basically be expressed as a financial benefit. Of course, in other cases, such as for public agencies, such a quantitative assessment is more difficult, in particular, the business model framework may not fit, or not directly, at least (Osterwalder & Pigneur 2010, p. 50). In such cases, however, one may find another scheme that at least allows for a qualitative assessment of the data value, such as the administrative tasks and competencies defined by public law for which the data is collected (cf. Konferenz der Unabhängigen Datenschutzbehörden des Bundes und der Länder 2019, pp. 37 et seq.). Nevertheless, in all these cases it is possible to determine a concrete qualitative or even quantitative value on the basis of the specific context and purpose of data use.

The value of data can also be determined qualitatively by the business model or administrative task for which it is used.

The same is true with respect to the risks, which can equally be determined on the basis of the specific context and purpose for which the data is used. The website visitors may see, for instance, in the setting of cookies for personalised advertising a violation of their privacy as well as a manipulation risk to their autonomous purchasing decisions. The owner of the website might see, in turn, the compliance risk of violating data protection law if not properly retrieving the website visitors' consent (cf. Krotova & Spiekermann 2020, p. 19). Further, the website owner may see a risk to reveal business secrets when sharing statistics about the visits on its website with competitors. In all these cases, it is also possible to assess the specific risks of data according to the specific context and purpose of data use. Thus, on the basis of a defined use case, assessing the value and risk of data does not pose a real problem (Ladley 2019, p. 34).

A slightly bigger problem would arise if there is no defined data use case. In such cases where the data use remains unspecific, the value and risk of the data remains equally abstract. Nevertheless, it is possible to assess the value and risk of data even if only in an abstract manner. In fact, most of the hype and concern around big data refers to this abstract value and risk (cf. Günther et al. 2017, pp. 195 and 197-198). To sum up the discussion, the general value of data lies in its machine readability, and therefore, with increasingly better technological capabilities, in the ability to process more and more data in different formats ever faster. In contrast, the concrete value depends on how the gathered information is used: for example, for what concrete business case or public task a private or public body uses the gathered information (Mayer-Schönberger and Cukier 2013). Corresponding to this, the bigger the data, the greater its general risk. Especially in data protection law, it is meanwhile possible to draw on a relatively elaborate risk assessment methodology. For instance, the greater the volume of data, the likelier the general risk is that the data might be misused; the more sensitive the data is, the more severe such misuse could be for the parties concerned. However, these risks remain abstract and vague as long as the data user does not specifically use or, at least, does not specifically intend to use the gathered information in such a way (that causes real harm or, at least, a specific risk against an individual concerned) (cf. von Grafenstein 2021b referring to Britz 2010; Albers 2012; Pombriant 2013). So even if the specific use cases remain undefined, it is possible to determine the value and risks of the data, albeit only in abstract terms. So far, this interplay of abstract and concrete value and risks of data has been recognised as a problem of how to preserve a potentially specific value in the long run or, respectively, how to prevent a potentially specific risk.[9]

Without a concrete use case, Big Data is only of abstract value

2.4.2 Falling apart of value propositions and risk expectations

However, a specific data governance problem arises, namely in data sharing, when concrete and abstract value propositions or concrete and abstract risk expectations fall apart among the stakeholders involved. This is especially the case if a stakeholder who has an interest in someone else's data has not yet been able to specify its own value creation to the extent that it could offer the data holder a corresponding concrete value proposition in exchange for the data. In view of known decision-making heuristics, this situation creates a problem because a stakeholder generally attaches less weight to an abstract value proposition than to an abstract or even specific risk (cf. regarding risk-aversion and endowment effects at Mousavi and Gigerenzer 2014). Since a data holder usually sees, at least, an abstract risk for itself in the disclosure of its data, the data sholder is likely to only exchange the data for a *concrete* value proposition. This might explain, for instance, the (alleged) privacy paradox (Müller et al. 2012, p. 175): Data subjects are often offered the concrete benefit of a free service in return for disclosing their personal data, while the associated risks are kept abstract and vague (which is clearly conflicting with the transparency requirements of data protection law if there is, in fact, a specific risk to the data subjects).[10] Therefore, against this backdrop, the behaviour of data subjects is not as paradoxical as it may seem (this *per se* rational calculus, is sometimes overlooked in the privacy calculus debate, see for example at Bélanger & James 2022, pp. 522 et seq.; however, cf. Acquisti et al. 2020, p. 742, as well as Riemensperger und Falk 2019, p.128).

Data holders only share their data voluntarily if this means getting a concrete value for running an only abstract risk.

In contrast, a data holder is much more unlikely to disclose its data if there is no specific value proposition in return for running into the abstract risks that the data disclosure causes. Such a lack of specific value propositions, however, is a rule of data-driven innovation (Spiekermann 2019, p. 209). Especially with exploratory big data analysis, the data user must usually first gain access to the data to find out what to concretely innovate. In the course of this innovation process, a value proposition may or may not materialise. However, the data holder has taken the abstract risk of data misuse by disclosing its data in any case. In many cases it is worse, as the data holder is exposed to pretty concrete risks.

Given the extremely broad scope of data protection law, for example, a lot of data may be personal data relating to the data holder's end customers. In this case, the data holder runs into a specific GDPR compliance risk. This is an astonishing result given that there may be hardly any concrete, but rather abstract, data protection risks for the data subjects themselves. This result is due to the fact that data protection law protects data subjects not only against specific but also abstract risks (von Grafenstein 2021b). With a regulatory approach that also protects against abstract risks and therefore results in an extremely broad scope of application, the legislator thus exacerbates the data-sharing dilemma by turning what is "only" an abstract risk for the individuals (protected by this law) into a concrete compliance risk for the regulation addressees. The situation is further complicated by the fact that the provisions of data protection law are so broadly defined due to their wide scope of application that the regulatory addressee can only calculate the compliance risk to a limited extent. The compliance risk is, therefore, perceived to be rather high in practice.

Data protection law turns abstract risks for data subjects into concrete compliance risks for data holders

Beside data protection law, the data may also contain trade secrets or fall under critical infrastructure security protection. In all these cases, a data holder runs specific risks by disclosing its data, which may not be worth it in light of an only abstract value proposition. In conclusion, if the data user cannot make a concrete value proposition to a data holder in exchange for at least the abstract risk that the data holder would incur by sharing the data, the data holder may not share the data. This is, indeed, a real data

governance problem (cf. European Commission 2020a, pp. 7 et seq.).

2.4.3 Reducing risks and maximising value to make data sharing worth it

To solve this problem, on the one hand, data holder and data user would have to (be able to) minimise the risks of such a level that the data holder would consider an even abstract value proposition to be worth that risk. For example, technologically, the data holder could use encryption measures to prevent other parties, who do not have the decryption key, from accessing the information. However, de facto, the data holder can hardly prevent a data user, with whom the data holder has entrusted the key, from passing on the decrypted data to further parties without the data holder's permission. By means of contractual agreements, the data holder could, therefore, also legally oblige the data user to not disclose the data to unauthorised persons, or at least to use it only for certain purposes. However, it is difficult to prove whether the data user actually complies with these requirements. Such control may only be possible again by certain technological means, such as intrusion detection systems or organisational measures, such as on-site physical control (cf. the use context-based approach promoted at Elliot et al. 2016, pp 52 et seq.). In this context, a data intermediary can again take on an important coordinating function, especially if the intermediary can credibly ensure and prove, as a trusted third-party, that the data user complies with the relevant protection rights. In data protection law, a certification body may perform such an accountability function according to Art. 42 et seq. GDPR (von Grafenstein 2021a, pp. 9-10).

On the other hand, data users may also seek to maximise the abstract value proposition for the data holders or as soon as possible turn the abstract value proposition into a specific value proposition. Perhaps this need for concrete value propositions is the reason for the claim for metrics to quantitatively assess the value of data. Especially with a price tag, there may indeed be an (additional) incentive for data holders to pass on data and take the risk in return for direct financial compensation. In my opinion, however, it is already a decisive step forward if the stakeholders involved make these specific value propositions at all, even if this happens only on a subjective basis in individual exchange relationships. Such a specific (subjective) value proposition can be a free service or even moral satisfaction (e.g. in the case of sharing data for research purposes, as in the health sector). In contrast, before an objective market price for data emerges (i.e. before a stable exchange of abstract risk expectations and concrete value propositions takes place between a sufficient number of data users and data holders), it will take some time – if it happens at all (Krotova & Spiekermann 2020, p. 30). Come what may, finding the right balance between risk expectations and value propositions in a constantly changing environment makes data sharing an extremely complex task.

Objective market prices for data will only emerge when there is a stable exchange of abstract risks (expectations) and concrete values (propositions) between a sufficient number of data users and data holders - this may take time.

In conclusion, one has to assess for which actor the data has what value and risk. On this basis, one can assess the necessary measures (i.e. legal, organisational or technological) to exploit the value and control the risk, respectively. Only on this basis, one can finally say whether the value is worth the risks and, not least, the costs of taking the measures to exploit the value and control the risks.[11]

2.5 Different degrees of centralisation

The last chapter focuses on a structural principle that can be described at each of the data governance layers as the degree of centralisation (Günther et al. 2017, p. 197; Abraham et al. 2019, p. 429). This structural principle is the focus here because the degree of its implementation in practice has a significant influence on

both the value creation and risk control as well as on the costs. Consequently, the principle is opening up a conflict of objectives.

With respect to the regulatory and organisational layer, an advantage of decentralised structures is that they make it easier to capture and reconcile different perspectives between the stakeholders involved on the value and risks of data. In IS literature, authors discuss, for example, the advantages of decentralised structures for the value creation through big data to communicate and involve different business stakeholders (Günther et al. 2017, p. 197). A similar thought is also discussed in literature on open innovation. In this context, the involvement of external stakeholders, such as (potential) partners, customers and/or consumers, is seen as an important prerequisite to bolster value creation (Chesbrough 2003).

Similar questions are discussed with respect to the degree of participation in the rule-making process in law and, more generally, with respect to political deliberative processes. Even though laws in civil liberties and societies are legitimised through democratic elections, there are numerous considerations here for incorporating further participatory elements into the concrete application of laws. The EU legislator, for instance, has enacted the General Data Protection Regulation for the protection of data subjects by formulating principles and rules for the processing of personal data. However, as its name suggests, the GDPR is just a *general* framework, which needs, therefore, to be specified in relation to the particularities of a specific case. The question of which actors set the principles and rules for the data processing in such a case, therefore, means *who interprets* and *specifies* the law? Interestingly, the GDPR barely mentions data subjects as stakeholders who actively interpret and specify these legal provisions. One of the few examples is Article 35 sect. 9 GDPR regarding the so-called data protection impact assessment, which recommends that the data controller shall seek the views of data subjects on the risks of its processing in question. Apart from this case, the dominant actors who specify and interpret the law are usually controllers (and their lawyers), data protection authorities and legal courts. This observation is even true with respect to the consent of data subjects. The reason for this is that the controller specifies the purpose of its data processing activities, and consequently, all further conditions that make the processing legally compliant (Art. 29 Data Protection Working Party 2013, p. 15). In contrast, the data subjects can only choose whether to accept such a preformulated consent or not. The only starting point for further participation of data subjects is the 'data protection by design' approach as established under Art. 25 sect. 1 GDPR. To guarantee the required *effectiveness* of the technological and organisational implementation of the legal requirements (e.g. transparency and consent), empirical design methods from HCI research strongly recommend to directly involve data subjects in the design process (von Grafenstein et al. in review). Other forms of direct participation can, of course, also be found in other areas of law, such as in municipal law (e.g. the referendum) or in public building law (e.g. public participation). From a political science perspective, in all these (and further) cases, the equal, inclusive and public involvement of the affected stakeholders is expected to lead to more appropriate decisions than just majority decisions (Habermas 1995; Beierle 2002; Friedman and Miles 2006).

Stakeholder involvement helps to identify the value and risk of data (use) as well as the appropriate exploitation and control measures

The degree of centralisation also plays an important role on the technological layer. In data protection law, for example, decentralised structures are considered to be less risky than centralised structures because an attacker from within or outside the involved entity/ies cannot simultaneously access all data at once if the data is stored and/or processed decentrally.[12] Also with regard to other risks, (e.g. the protection of trade secrets or a competitive disadvantage resulting from the disclosure of data), the actors involved in data

sharing tend to prefer decentralised solutions. According to our observations in such situations, decentralised structures enable the involved entities to maintain control over who they exchange data with and under what conditions, rather than if all data for all possible uses were collected centrally in advance and all access and use conditions were set by a single body, e.g. the state or a market-dominant private company (cf. the single-sign on solution https://netid.de/, which avoids that the partnering companies – which are competing against each other on the media content market – share only as much data as necessary to get their customers' consent more easily).

However, such decentralised mechanisms also have their disadvantages. First of all, more decentralised and participatory structures undoubtedly lead to an increase in the complexity of the corresponding processes. It is obvious how much more complex the processing of personal data becomes when data subjects are actively involved in the technological and organisational design. Decentralised structures may also conflict with the so-called principle of congruence from organisational theory, which states that tasks, responsibilities and competencies should coincide in order to ensure goal-oriented actions within a company (Otto 2011a). Furthermore, decentralised structures on the technological level may also hamper data quality and IT security (Günther et al 2017, p. 197, with further references). Although the disruptive potential of decentralised IT structures has been recognised, at least, in certain fields (see, for example, the excitement around the blockchain), the discussions about decentralised server solutions (Gaia-X), distributed computing (Kahanwal & Singh 2012), in particular local privacy solutions such as local differential privacy (Ye & Hu 2020), show that all this is far from being established as a market standard.

However, decentralised structures and processes also increase the complexity of data governance

Against this background, it is interesting to assess which degree of centralisation is best suited to which type of conflict of interest (Otto 2011b, pp. 60 et seq.). Contreras and Reichmann make an important step toward such an assessment with respect to the sharing of scientific data between entities. In their work they observe the following "four basic structural models (...) along a continuum ranging from the most to the least centralised (see the table).

> (i) *fully centralised*: all data is aggregated in a single, centrally managed repository;
>
> (ii) *intermediate distributed*: repositories are distributed and separately maintained, but may be interconnected by a central access portal, share technical service components, and utilise a common data-exchange format [sometimes called a federated database system];
>
> (iii) *fully distributed*: repositories are maintained locally and are not technically integrated, but share a common legal and policy framework that allows access on uniform terms and conditions (legal interoperability);
>
> (iv) *noncommons*: repositories are largely disaggregated and lack technical and legal interoperability and, at most, may share a common index." (Contreras & Reichman 2015)

Both authors evaluate the advantages and disadvantages of these observed patterns: While they recognise the fully centralised models positively in terms of better data quality, (but also negatively in terms of their higher costs), they are deterred by the noncommons because of their complete lack of interoperability. As a consequence, they highlight the models in between that provide technical and/or legal interoperability but at lower costs than fully centralised models (Contreras and Reichman 2015).

Therefore, one of the key data governance issues is over which centralised or decentralised structures, at the regulatory, organisational and technological level, are best suited to resolving the conflicts of interest within and between entities. Intermediaries will play an essential role in this process as their function makes them most likely to be able to generate and maintain the necessary knowledge, structures and processes to resolve these conflicts on a large scale. Thus, in addition to the aforementioned coordination and trust function, another essential function of intermediaries is to reduce costs through economies of scale. Another publication has already taken up the discussions on the function of intermediaries in IP governance to develop initial models for data governance intermediaries (Wernick et al. 2020, p. 67). Building on the present framework, these models now need to be specified in such a way that they can solve the challenges for data governance described here, both sector-specifically and across sectors.

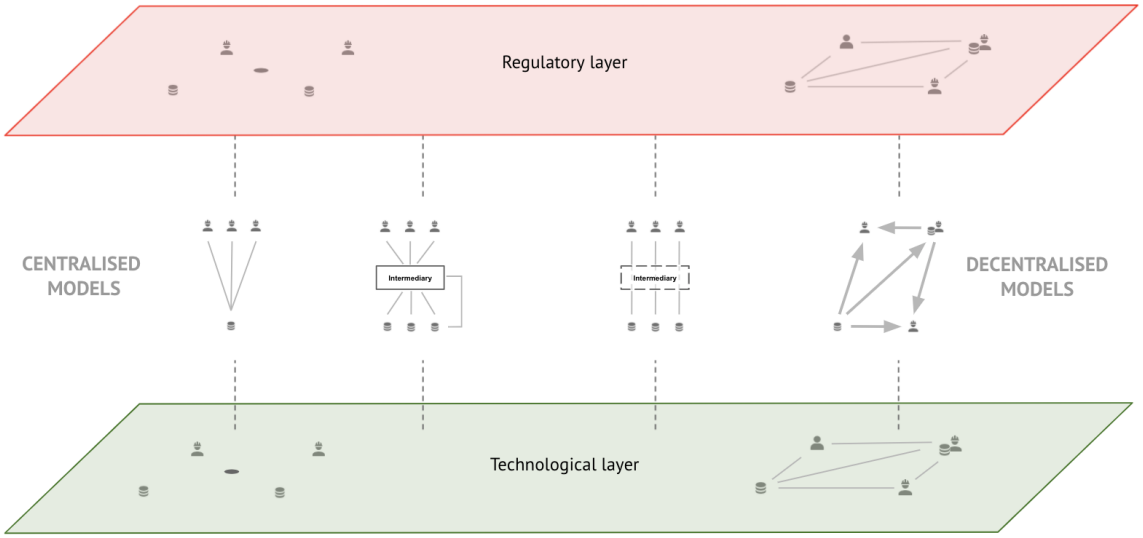

Fig. 2: Data governance models with different degrees of centralization (following Wernick et al. 2020, p. 67)

3. OUTLOOK: HOW TO IMPROVE DATA GOVERNANCE?

This contribution criticises the ambiguity of several key terms and concepts currently discussed in IS literature and aims at elaborating on a refined framework to better describe certain challenges and conflicts in the field of data governance that arise especially on highly regulated markets. The refined framework shall enable further research to examine and compare, more precisely, the success factors of potential data governance solutions. On this basis, further data governance research, whether still conceptual or empirical, may enable more effective public and private, evidence-based regulation.

3.1 Summary of the proposed data governance framework

To reach this aim, the paper argues that the main goal of successful data governance is to reconcile conflicting interests in data (use, reuse, sharing, etc.). A major challenge here is that the involved actors must coordinate their different perspectives on the value and risks of data, which can change continuously depending on the respective purpose and context of data use while cooperating on different governance

layers, (i.e. the regulatory, organisational and technological layer). This leads to a "value for risk" dilemma, especially concerning data sharing in the highly regulated EU Single Market: While data owners often see a concrete (compliance) risk in the disclosure of the data, users can usually only give them an abstract value proposition for the sharing of the data if they have not yet been able to specify the use case, as is often the case with data-driven innovation. For reasons of behavioural economics, a concrete risk is rarely exchanged for an abstract value proposition. Accordingly, data holders in such situations tend not to share the data. So as long as data is to be shared voluntarily or, at least, regulation leaves room for interpretation and manoeuvre, all the actors involved must constantly coordinate who takes which necessary legal, technological and/or organisational measures so that data sharing is worthwhile in view of the (in the best case now maximised) value of the data and the (now optimally controlled) risks. To find the right point where the expected value exceeds the risks and costs, the degree of centralisation on all three data governance layers is a decisive element. This refined framework may, for example, help to flesh out the details of the European data spaces (European Commission 2020a).

3.2 Conclusions for some Digital Services Package draft laws

In finding the right balance between value creation and risk control, in short whether and how data is shared, the regulator plays a crucial role. This is especially true for the EU Single Market. With every protection law, the regulator may contribute to the synchronisation of the possibly different risk perceptions among the actors involved. However, from the perspective of a data holder or data user, each protection law applicable to the collection, reuse and/or sharing of data represents a further concrete compliance risk. This, in turn, is often matched by an unclear or abstract value proposition. Thus, if regulators do not want to stifle but rather encourage the reuse and sharing of data, they must take appropriate countermeasures for each protection law that helps overcome the "value for risk" dilemma that the law itself creates. A tried and tested means of doing this are data sharing duties and access rights, as well as all kinds of clearing centres (in its broadest sense) including legal conformity assessments (cf. Wernick et al. 2020). Interestingly, in its digital services package, the legislator resorts to various of these mechanisms, some of which will be briefly analysed with respect to their suitability in solving the value-for-risk dilemma as described. When analysing the legislative package, it becomes clear that the legislator has proceeded phenomenologically (from the regulation of market power to platforms to AI systems to data governance, etc.) and has carried out extensive multi-stakeholder procedures in each case in order to record the respective value and risk perceptions of as many potentially affected parties as possible and to collect possible measures to solve these conflicts of interest. However, it is obvious that the laws have not yet been consistently coordinated with each other (see already Graef & Gellert 2021) and that some laws in themselves are, in a regulatory ratio, not yet sufficiently precise.

3.2.1 Opening / accessing protected data held by public sector bodies (Chapter II Data Governance Act proposal)

Let us start with opening and accessing data held by public bodies. The second chapter of the Data Governance Act proposal extends the Open Data Directive, which only concerns the publication of data held by public bodies as long as the data is not covered by protection laws, such as on the grounds of commercial or statistical confidentiality, intellectual property or data protection laws. To this aim, the legislator harmonises in its Data Governance Act proposal certain (legal, organisational and technological) conditions under which public bodies may publish certain data that is protected by those rights. First of all, the proposal requires the EU Member States to install a single information point where data users can find all available data ressources, the conditions for its re-use as well as the corresponding fees (Art. 5, 6 and 8).

Through this single information point, public bodies may grant or reject access requests by data users. Furthermore, data users have the right to contest such a decision before the courts of the respective Member State.

This proposal definitely has the potential to reduce the coordination costs for the actors involved not only by harmonising a legal, technological and organisational framework for the sharing of data held by public bodies, but also by centralising the related information through a single access point (see the intermediate distributed solution mentioned above by referring to Contreras and Reichman 2015). However, it only has the potential because there is no obligation of public bodies and certainly no data access rights for data users. Whether and to what extent public bodies make such protected data available through the corresponding single access point depends on the free decision of the public bodies. Thus, without a legal right to data access, the applicants' right to contest such a decision is rather feeble.

In contrast, the legislator should at least provide for an obligation of public bodies to catalogue all protected data held by them and to make these catalogues accessible via the single access point. In doing so, the public bodies may meet the protection interests of natural or legal persons, since these data catalogues do not contain the raw data but only metadata (i.e. descriptions of the raw data), which can be specified in different degrees of aggregation or abstraction, depending on the need for protection. Such data catalogues are important, because only such catalogues enable potential applicants to obtain an overview of the data that is basically available and to submit corresponding access requests. On this basis should even further steps be implemented to support the sharing of data more effectively: At first, on the basis of these (more informed) access requests, the authorities can (more reliably) systematise, evaluate and prioritise the respective processing purposes (which would have to be specified in the request form) with regard to their frequency, practical relevance and need for technological and organisational safeguards. On top of that, more serious thought should be given to providing applicants a general data access right, but on the condition that the relevant protection laws can be met. This access right should be complemented by a right of the applicant to propose and introduce alternative technical and/or organisational measures (at their own expense and, if desired, through a third party) if, in the applicant's view, these measures are equally protective for the case in question and better exploit the usefulness of the data than those initially required by the authority. This would significantly strengthen the right to a judicial remedy by extending it to the question under which measures the right of access exists and which measures are appropriate. This extension also has the positive side effect of increasing the number of court decisions that would counteract the high degree of legal uncertainty in this area (especially in data protection law) that has resulted from the high enforcement deficit to date.

Private bodies should get an access right under the condition that all protection laws are met, while public bodies should be obliged to catalogue their data

In this context, four aspects should additionally be clarified in brief: First, technological and organisational measures, such as anonymisation and pseudonymisation, can not only be used to protect personal data but also other sensitive information, such as trade secrets (see, in contrast, the misleading wording in Art. 5 sect. 3 Data Governance Act proposal; Bitkom 2020, p.44). Second, the legislative indication that the protection measures can, depending on the need for protection, exist in a secure digital processing environment provided and controlled by the public body or even only within its physical premises ("where the secure processing environment is located"), is an important clarification emphasising the need for taking the circumstances in which the data is used into account (see Art. 5 sect. 4 Data Governance Act proposal; and already above with reference to Elliot et al. 2016). Thirdly, there is a special case in data protection law with regard to the legal basis that is required in addition to the protection measures. Art. 5 sect. 6 Data

Governance Act proposal gives the impression that the last resort is to assist applicants in obtaining new consent from the data subjects if no legal basis from the GDPR is applicable. In fact, there is always the option of creating a new legal basis pursuant to Art. 6 sect. 1 lit. e and sect. 3 GDPR, whereby this legal basis must specify, among other things, the concrete purposes of the data processing and the protection measures. On the basis of a systematised analysis of the access requests (and the processing purposes stated therein), the public body can therefore also suggest to the competent legislative body the creation of corresponding legal bases. Thus, in addition to technical and organisational support, the so-called competent bodies according to Art. 7 Data Governance Act proposal can also assist with systematically prepared proposals for these legal measures. Last but not least, it seems obvious to designate the national data protection authorities as "competent bodies" due to their historically renowned expertise on the interdependence of legal, technical and organisational measures (see point 2.3.2 above). Of course, this presupposes a corresponding (significant) increase in resources so that the authorities can proactively conduct the corresponding analyses and develop and provide for appropriate technical and organisational measures.

Data protection authorities should be entitled and obliged to analyse, develop and suggest concrete solutions to best exploit the value and mitigate the risks

Of course, with such obligations on public bodies and corresponding access rights for applicants, the public sector primarily bears the costs of successfully coordinating and resolving the conflicts of interests in the respective data. The Data Governance Act proposal provides insofar that the costs must be non-discriminatory, proportionate and objectively justified and shall not restrict competition (Art. 6 sect. 2). For refinancing reasons, however, it would be worth considering making the amount of the fees dependent on the financial capacity of the applicant and abandoning the prohibition of discrimination not only in favour of SMEs (cf. Art. 7 sect. 4) but also at the expense of at least so-called gatekeepers in the sense of the Digital Market Act proposal (cf. the similar differentiation in the Data Act draft excluding SMEs from sharing duties and excluding gatekeepers from accessing the data).

3.2.2 Accessing data held by private parties (esp. Chapter II and V Data Act proposal)

The opposite case, where public bodies wish to access data from a private body, is dealt with in chapter V of the Data Act proposal. These provisions are designed as a kind of European "backup" access right: According to Art. 15 Data Act draft, at first, a private data holder must only grant a public body access to its data if the data requested is "necessary to *respond* to a public emergency" (lit. a – italics added by the author). However, second, the sharing obligation also exists if the data is necessary to "*prevent* a public emergency or to *assist the recovery* from a public emergency" (lit. b – italics added by the author). And finally, the obligation even exists in case the public body needs the data to "fulfil a specific task in the public interest that has been explicitly provided by law" and the public body "has been unable to obtain such data by alternative means, including by purchasing the data" or (!) "by relying on existing obligations to make data available, and the adoption of new legislative measures cannot ensure the timely availability of the data" (lit. c). This means in brief, as long as the national legislator does not get its act together to create a data sharing obligation in time, the authority can invoke the Data Act provisions. It is worth mentioning in this regard that these access rights are excluded for certain purposes, especially in connection with criminal penalties, customs and taxes (Art. 16 sect. 2). The public bodies are also subject to a kind of principle of purpose limitation (Art. 19 and 21) – as in data protection law. This means that the public bodies must delete the data once the purpose has been achieved and may not use it until then for any other purpose that is "incompatible" with the purpose for which the

A European "backup" access right for public bodies

authority accessed the data (however, as we know from data protection law, there is room for manoeuvre, see von Grafenstein 2020b, 2021b, 2021c). Last but not least, the authorities must implement appropriate measures to protect personal data and trade secrets (Art. 19) and bear the related costs, including a reasonable margin. However, the data holder must explain the basis for its calculation in sufficient detail to allow verification of the costs and reasonable margin (Art. 20).

In contrast, sharing duties or access rights between private bodies only exist for data that is generated through the use of a product or related service, and only for the benefit of the respective user (Art. 4) or if the transfer to a third party is at least mediated by that user (Art. 5). In this sharing constellation, the access rights and sharing obligations are again subject to the (mainly clarifying) condition that this must be done in compliance with the GDPR and the protection of trade secrets (Art. 4 no. 3 and 5 as well as Art. 5 no. 6 and 8). For these sharing constellations, the duty to implement the necessary organisational and technological measures arises in part directly from the legal draft itself (Art. 5 sect. 8) and in part only by reference to the other applicable law, in this case the GDPR (Art. 5 sect. 6). In the second case, however, the legislator should definitely clarify that the reference in Art. 5 sect. 6 Data Act draft to the legal bases under Art. 6 and 9 GDPR is only of an emphasising nature. It is not intended to exclude all the rest of the GDPR provisions. To avoid misunderstandings, the reference should therefore be to all provisions of the GDPR, in particular to the approaches of data protection by design and security of processing (Art. 25 and 32 GDPR). If Art. 25 and 32 GDPR did not apply, there would be no technical and organisational protection against the data protection risks for fundamental rights caused by private data sharing.

More interesting than this clarifying reference, however, is the additional condition that the actors involved may not use the data received for the development of products that compete with the other actor's products or affect its economic situation (Art. 4 nos. 4 and 6, Art. 5 no. 5 and Art. 6 para. 2 lit. e). This is interesting because this interest is not necessarily covered by the already existing trade secret protection. Furthermore, if a third party (data user) receives data via the user of a product or related service, this third party may only use the data for the purposes specified by the product or service user – similar to data protection law – and in particular may not pass the data on to other users (Art. 6 sect. 1 and 2 lit. c). Even more interesting are the sanctions provided if a third party (data user) does not comply with these conditions. In this case, the data user must not only destroy the data provided by the data owner and any copies thereof, but also cease the production or use of goods, derived data or services produced on the basis of knowledge gained through this data (Art. 11 para. 2). This second sanction does not apply only if the use of the data has not caused significant harm to the data holder or if it would be disproportionate to the interests of the data holder (Art. 11 para. 3). The most interesting question here is, of course, what requirements are placed on the evidence with which a data holder must justify the accusation of a violation of these requirements or with which a data user can exonerate itself from such an accusation. The interest here stems from the fact that misuse of information is usually difficult to prove (von Grafenstein 2021b). Since the law does not specify any further requirements in this regard, this question will probably be clarified, to a limited extent, by the dispute settlement bodies (Art. 10) and, more comprehensively, by the competent authorities (Art. 36) before which the stakeholders can have their disputes clarified.

It is worth mentioning that these sharing obligations and access rights depend, furthermore, on the market power of the actors involved: SMEs are not obliged to share data generated by its products or related services (Art. 7 sect. 1) and gatekeepers are excluded from the group of third parties with whom the data may be shared (Art. 5 sect. 2). The Digital Markets Act contains further access rights and restrictions with regard to gatekeepers. For example, according to the version consolidated in the trilogue, gatekeepers must "give sellers access to their marketing or advertising performance data on the platform"; at the same time,

gatekeepers may "no longer reuse private data collected during a service for the purposes of another service" (EU Council 2022). SMEs are also privileged in terms of costs since the compensation they have to pay to a data holder for getting access to data "shall not exceed the costs directly related to making the data available" (Art. 9 sect. 2). In all other cases, the costs must at least be reasonable (Art. 9 sect. 1). Here again, the data holder must explain the basis for the calculation of compensation in sufficient detail to allow verification of compliance with the cost requirements (Art. 9 sect. 4). The costs will be another issue that will be answered by the dispute settlement bodies and competent authorities. By the way, in addition to the costs of the data holder, the data user will also have to bear its own costs resulting from the implementation of the technological and organisational measures on its part. Whether these costs are worth it to gain access to the data will be decided on a case-by-case basis over a certain period of time.

Access rights and sharing obligations for usage data (as well as costs) depend on the size and market power of the private companies

As mentioned before, the potential for clarifying the numerous legal issues that arise when private data users access data held by public bodies is still under-exploited in the current draft of the Data Governance Act (however, see the proposed approach above). In contrast, the current draft of the Data Act applies a far more comprehensive approach that enables the stakeholders involved (or interested) in the sharing of data to clarify many legal issues that may potentially arise. This is more than welcome given the high legal uncertainty in this area and will not only help the actors involved in a concrete conflict in the short term, but will even more so help the data-driven economy in the long term as a whole. A necessary condition for the systematic analysis of these conflicts and solution measures is that the conflicts are brought before suitable dispute settlement bodies or competent authorities (cf. the cooperation mechanism in Art. 32 sect. 3 Data Act proposal). Of course, these bodies must then also be provided with the sufficient resources, especially in view of the complexity that the aforementioned conflicts of interest and interdependencies of the solution measures on the legal, organisational and technological layers entail.

In conclusion, it is interesting to ask, given the aforementioned conditions are met, whether the scope could not have been extended to a general data access right? If access to data is only permitted if all legally protected interests are respected, such as data protection and trade secret protection, and even the economic interest of data holders does not suffer competitive disadvantage or unwanted competition due to the disclosure of its data, why not provide for a general data access right? Why shouldn't any private data user have the right to approach any private data holder and request access to their data if the above conditions are met and the data user bears the reasonable costs? The legislator did presumably avoid such a radical step since this would be a far-reaching interference in the fundamental rights of the data holders, despite the comprehensive weighing of all possible opposing interests. In the current draft of the Data Act, the legislator considers such an interference to be justified not only for the general reason of bolstering the data-driven economy but also because this ultimately strengthens the individual "right to use and dispose of lawfully acquired possessions" through a right of access to data generated by the use of such a possession (see the Memorandum, pp. 7-12). The exact legal-political or even legal-philosophical justification of such an ownership idea is another matter; in any case, this justification does not apply to a general access right. For this reason, the legislator relies more on the voluntary nature of actors involved in these further data sharing constellations and rather helps them to overcome the value-for-risk dilemma through a variety of softer legal measures.

There is no general access right for private parties (going beyond usage data), instead, voluntary data sharing is supported through a variety of "soft laws"

Before we come to these soft law measures, it should be asked under which circumstances the introduction of a general access right would be justifiable. On closer examination, the introduction of such a general data access right only makes sense if at least the following three conditions are met: a) no structures or processes have yet emerged between private parties according to which the data sharing dilemma could be satisfactorily overcome on a voluntary basis (i.e. insufficient corresponding risk minimisation measures or value realisation mechanisms); b) the competent authorities have succeeded in systematising conflicts of interest that typically arise with regard to usage data and in developing suitable solution measures; c) it was also possible to clarify whether and how successfully an allegedly infringed party can prove a misuse of data with the consequence that the opposing party must destroy its products resulting from the proven misuse of data. Since it takes some time until all three conditions are actually fulfilled, the data access rights introduced in the current draft laws should be used as a kind of testbed, waiting about 5 years to see whether the structures mentioned in connection with these access rights emerge. In order not to have to start and carry out a complete legislative procedure for the introduction of the general access right in such a positive case, the general access right should already be introduced now, but under a moratorium of 5 years. Only if all three conditions are not (yet) met can the application of the general access right be further postponed. In any case, the commission must make its own decision, which it is obliged to do in the present law.

A general access right should be introduced under a moratorium of about 5 years to ensure that the necessary structures are in place.

3.2.3 Defining and promoting data intermediaries more effectively (esp. Chapter III and IV Data Governance Act proposal)

A central starting point for such soft legal measures is the establishment, harmonisation and support of data intermediation "services" (in its broadest sense). As explained above, there are several reasons why data intermediation services in general could significantly help to overcome the described data sharing dilemma: Due to their independent role and focus on intermediation between potentially many stakeholders, data intermediaries are possibly best positioned to generate knowledge about the conflicts of interests and design the appropriate mechanisms to solve them (i.e. how to exploit the value and mitigate the risks), whilst also benefiting from economies of scale (irrespective of whether these economies of scales serve profit or non-profit goals). In this respect, they also form an important counterpart when dealing with the (possibly restrictive) solution practices of the competent authorities. If these authorities provide solutions with technical and organisational protection measures that disproportionately reduce the value of the data from the point of view of a data user, the data user can challenge this decision. Data intermediation services can provide important support here due to their specialised knowledge. In any case, data intermediaries can fulfil this coordinating function in many different roles at all three governance levels, whether in the form of individual mechanisms at single layers only or in the form of fully integrated services. In this regard, it is worth emphasising that intermediation services can not only help the actors involved in data sharing to overcome the value-for-risk dilemma in voluntary sharing constellations. Intermediation services can also provide the technical and/or organisational protection measures that data holders must fulfil in order to exercise their access rights.

Data intermediation can fulfill its coordinating function in different forms at all three data governance layers, whether as single mechanism at one specific layer (e.g. an anonymisation service) or in the form of fully integrated services

Examples of intermediation services on a technical-organisational layer are anonymisation and pseudonymisation services offered by private companies, or on a legal layer the harmonisation of

contractual sharing terms (see the model contractual terms to be developed by the EU Commission, Art. 34 Data Act proposal), or on a more legal-technical (interface design) layer, the harmonisation of consent forms (see the European data altruism consent form in Art. 22 Data Governance Act proposal). Indeed, such standardised contract or consent mechanisms can take on a considerable degree of complexity including the technological and organisational layers, depending on the extent to which the specifications for obtaining, storing, forwarding and revoking such contracts or consents are standardised or even made available as a technical-organisational infrastructure (which in turn can follow a centralised or decentralised architecture, see already under point 2.5). In principle, this would be possible for all kinds of areas and not only non-altruistic areas (see Art. 22 Data Governance Act proposal focusing on altruistic forms so far). As long as this would continue to be an offer, only, allowing other actors to further develop evidence-based state of the art solutions (see, for instance, the methodology for assessing the constantly evolving state of the art of GDPR transparency and consent measures, von Grafenstein et al. in review), this would indeed promote rather than hinder the scope for innovation in these areas. A last example mentioned here refers to intermediation services on an organisational layer, such as for organising bargaining power for the negotiation of data sharing terms (see the so-called data cooperatives under Art. 9 sect. 1 lit. c) Data Governance Act proposal).

The legislator should provide for as many data intermediation forms as possible, as long as these forms are non-exclusive and non-mandatory (leaving room for innovations)

In the following, however, the focus will be on services that integrate to an extent more or less all three data governance layers, as so-called data sharing services (for profit). This is where the greatest ambiguities exist in the Data Governance Act proposal. With its provisions regarding data sharing services, the legislator means "to increase trust in sharing personal and non-personal data and lower transaction costs linked to B2B and C2B data sharing" (see the Explanatory Memorandum in point 5). To reach this aim, the draft establishes, in essence, a notification requirement for such sharing services as well as a number of conditions, in particular what these services are not allowed to do. Above all, data sharing services may not use the *data* for any purpose other than making it available to data users, and may only use the *metadata* but only for the development of this service. Most of the other provisions are actually, here again, clarifications of regulations that are applicable anyway, above all those of data protection law and competition law (Art. 9 sct. 2 DGA proposal). Even with regard to (only relative) trade secret protection, the draft exhausts itself in a mainly clarifying function, since a data holder will usually pass on the technical, legal and organisational trade secret protection measures to the sharing service and, ultimately, to the data users on its own initiative, i.e. without the need for further legal support (see already point 2.3.2; however, see also Art. 11 no. 7 and 8 Data Governance Act draft on such measures for non-personal data). This raises the question as to whether the trust signal that is supposed to be triggered by the mere obligation of a data sharing service to register is strong enough to outweigh the aforementioned restrictions. Doubts that the current draft does not effectively support data sharing services, especially in overcoming the data sharing dilemma as described are fed by at least three reasons:

First, at a closer look, it seems overstretched to oblige data sharing services to use data exclusively for the sharing of the data (see Art. 11 no. 1 of the draft of the Data Governance Act – or even without any personal economic interest in the data, see Art. 26 sect. 1 no. 2 of the draft of the German Law on Privacy in Telecommunication and Telemedia). In principle, it is possible for intermediaries to find business models in which they also use the data for their own purposes, but in a way that does not involve any undue risks for the data holders (including data subjects) and data users. If an intermediary cannot demonstrate such a *low-risk* use for its own purposes, hardly any data holder or data user will use the service because the (compliance) risk of using such an intermediary is just too high for them. A legal regulation should

therefore be limited to oblige a sharing service to make transparent such risks that its own processing purpose poses for the data holders and data users who are involved or concerned by the data sharing.[13] In contrast, a further restriction, which goes beyond such a self-regulatory market selection mechanism and obliges the intermediary to use the data solely for the sharing service, would considerably limit its own innovation process in finding a functioning business model. This, in turn, would cut off the option to offer the sharing service for a lower price, or even for free. This conclusion also applies to the case that the underlying aim of the current regulatory draft is to actually exclude market-dominant companies such as Google and the like from acting as a sharing service. Would the legislator seek to eliminate those market power risks, the legislator should make this explicit and, for example, link it to the gatekeeper concept of the Digital Market Act draft, instead of also depriving SMEs of their innovation opportunities. Of course, the above transparency obligations would still have to apply to SMEs as well.

Second, as long as the notification duty of sharing services does not mean that the competent authorities also assess and clear the compliance of these services with applicable protection laws, this mechanism barely reduces the compliance risk of the actors involved in using and providing the services. In this case, the notification duty is only a weak, if not a symbolic, trust signal (which should not be underestimated in terms of its effect on markets, of course). However, it would be more effective to oblige data sharing services, as far as personal data is concerned, to adhere to certification mechanisms or codes of conduct in accordance with Art. 40 et seq. GDPR. Sharing services may use such mechanisms (even if only as an "element") to demonstrate GDPR compliance (Art. 24 sect. 1 and 3 and Art. 25 sect. 3). In the GDPR, such adherence is voluntary. However, to take into account the increased need of data holders and data users for legal certainty when using a sharing service, the Data Governance Act could make these services subject to an obligation (cf. Blankertz and Specht 2021). Furthermore, the legislator may establish similar mechanisms with respect to other protection laws, such as trade secret protection and competition law. For this, the legislator had to establish legal provisions according to which adherence to a corresponding code of conduct or certification programme can be used as proof of compliance with these other protection laws as well. In doing so, the legislator should also require the owner of such certification mechanisms or codes of conduct to specify its criteria (under which legal compliance is met) as concrete as possible: The more concrete these criteria, the better the competent authorities can verify the correct specification of the protection laws by these criteria in advance as part of the accreditation process.[14] In contrast, general criteria run the risk of creating black boxes that only conceal the high degree of legal uncertainty and further increase the risk of inconsistent legal interpretation (von Grafenstein, 2021d). This clarification was overlooked by the GDPR legislator and is now leading to corresponding problems in the interpretation and application of this law. Instead, a clear duty to adhere to concrete certification mechanisms or codes of conduct would decrease the compliance risk of the stakeholders involved in data sharing and, thus, expand the trust-building and mediating function of data sharing services significantly.

Instead of forbidding sharing services to drive their own data-based innovations, they should only be obliged to inform about the corresponding risks and adhere to a third-party conformity assessment

This leads us to the third reason why the current draft is questionable not only in its hoped-for positive effect on data sharing services, but also in its legal justification: the fairly vague scope of application. Seeking to clarify the scope, Recital 22 Data Governance Act draft states that such sharing services must be, amongst other aspects, "independent from both data holders and data users" and "assist both parties in a transaction of data assets between the two". According to this understanding, for instance, providers of cloud services should be excluded, as well as advertisement and data brokers, data consultancies, providers of data products that result from value added to the data, and data exchange platforms that are exclusively used

by one data holder to enable the use of their own data. Even if the exclusion of these examples from the scope of application is intuitively comprehensible, at second glance, the inherent connection remains opaque. This ambiguity becomes particularly apparent when, on the one hand, the exclusion of such service providers is justified by the fact that they aggregate, enrich or transform the data in a way that a direct relationship between data owners and data users is no longer established. On the other hand, data sharing that transforms the data in a way that improves their usability for data users is still covered by the regulation. How can one reliably draw a line between these two sharing services?

In my opinion, hardly at all. Instead, one could significantly sharpen not only the scope of application of these provisions but also their legal ratio if – as proposed above – the sharing provider is allowed to also transfer the data for its own purposes (and interests). In this case, it is possible to make a clear distinction with recourse to the definitions from data protection law (see already above at point 2.2.1): If the data sharing service provider processes the data (also) for its own purposes, the provider is a "controller" and thus (together with the data holder and data user) responsible for the processing (as so-called joint controllers). In contrast, if the provider processes the data exclusively for the purposes of the data holder and/or data user, then the provider is a "processor" with limited responsibility (mainly for maintaining IT security, Art. 32 GDPR). As long as the provider transfers the data exclusively on behalf of the data holder or data user (such as a cloud provider), the data holder or data user can fully control the provider's actions, so that there is no need for legal protection that should add to the GDPR. In contrast, if the provider (also) transfers the data for its own purposes, the data holder and data user do not have full control over the sharing provider, so that the resulting need for protection justifies additional regulation. Therefore, this definition means, for example, that the aforementioned data brokers fall within the scope of application because they help sharing data also for their own (commercial trading) purposes, even if there might be an only loose connection between initial data holders and data users. However, it should be clear for this last example that the law does not prohibit these data brokers from their activities. Rather, data brokers may process the data for their own purposes. The only condition they have is to apply special transparency obligations and submit to a certification mechanism or code of conduct. In view of the current practice of data brokers, which often takes place in a legal dark-grey area of the GDPR, this would be a real benefit for not only data subjects but also for the entire (online advertising) market, while the regulatory burden remains relatively low.

A notification obligation of data sharing services should be limited to sharing services acting as a "controller" (in the meaning of the GDPR)

These considerations equally apply to the sharing of non-personal data (see already above point 2.2.1) as well as to not-for-profit sharing services. The latter statement means that so-called data altruism organisations only have to register under Art. 15 Data Governance Act proposal if they also transmit the data for their own purposes. In contrast, if the sharing takes place solely on behalf of the data holder, there is no particular need for transparency on the part of the data holder (since the data holder fully controls its processor). However, if a sharing service collects and transfers the data solely on behalf of a data user, which is a not-for-profit entity and processes the data for charitable purposes, such a data user should equally be able to have this charitable status entered in the register. The reason for this is that such data users have the same need to demonstrate its charitable status and purposes as charitable sharing services who process the data for their own purposes. The legislator should, therefore, open up the possibility of registration under Art 15 Data Governance Act for not-for-profit data users. In conclusion, the registration duties for so-called data sharing services are less about a symbolic trust signal per se. Instead, the legislator should align its regulations more stringently with the concrete needs of the actors involved in data sharing: This need for a verified trust signal exists, on the one hand, for data holders and data users when they use a sharing service

that they cannot fully control due to a lack of full instruction authority. On the other hand, the need exists for sharing services and data users equally when they want to prove their special trustworthiness as non-profit entities to data holders.

3.2.4 Aligning the AI Regulation with the GDPR and Data Governance Act

Another legal means to overcome the value-for-risk dilemma described above are clearing centres, especially when building on conformity assessments. As explained before, the sharing of data gives rise, due to the multitude of applicable protection laws, to an even greater number of legal questions. These legal questions pose pretty concrete compliance risks to data holders as well as data users (and even for specialised intermediation services), while the value of the data often remains abstract and vague. However, such legal questions may either be clarified and, thus, the compliance risk be reduced through directly state-led procedures, such as by appealable competent authorities and judicial courts (Art. 8 sect. 4 Data Governance Act draft and Art. 36 Data Act draft), or by way of co-regulatory procedures, e.g. through dispute settlement procedures (Art. 10 Data Act draft), certification mechanisms or codes of conduct (Art. 40 et seq. GDPR; Art. 69 AI Regulation draft regarding codes of conduct for non-high risk AI systems). Another co-regulation mechanism are the so-called conformity assessments as foreseen in the AI Regulation draft.

The AI regulation draft plays a role here, not only because this regulation will apply as another protection law to the use, reuse and sharing of data, but also because the regulation will have a vast area of overlap with the GDPR (if not significantly changed during the legislation process). This overlap raises the question of how data holders and data users may counter the compliance risk resulting from these laws, first on a legal layer, but then also on an organisational and technological layer (see above point 2.3). The vast overlap results, on the one hand, from the extremely broad definition of AI in Annex I, which even covers statistical approaches and, on the other hand, from the fact that the so-called high risk AI-systems covered by the scope of the legal draft are defined for areas in which mainly personal data is processed in the AI systems (Annex III AI Regulation draft).This is mostly true for the following areas: Biometric identification and categorisation of natural persons; education and vocational training; employment, workers management and access to self-employment; access to and enjoyment of essential private services and public services and benefits; law enforcement; migration, asylum and border control management; and administration of justice and democratic processes (since the facts and applicable laws of a specific case interpreted by an AI system refer to natural persons involved in that case). Only in the area of critical infrastructures is no personal data necessarily processed in the AI systems. In this area, the regulation would therefore have a really autonomous scope of application. In the other areas, in contrast, data protection law is typically applicable, even if not all of them fall under the GDPR but under specific data protection laws that may take precedence over the GDPR (e.g. the EU Law Enforcement Directive). However, since these more specific laws often follow the same basic principles as the GDPR (e.g. the data protection by design approach in Art. 20 of the Law Enforcement Directive), the differences will not be discussed in detail here.

The AI Regulation draft and the GDPR vastly overlap, both laws also apply a risk-based approach to fundamental rights. It is lamentable that the AI Regulation draft does not look at more systemic risks (such as environmental protection) that are not already addressed by the GDPR's individual fundamental rights protections.

If one compares both laws, the GDPR and the AI Regulation draft in their regulatory approach, both are – at the latest and at second glance – astoundingly similar: Both laws focus on the purpose for which the personal data is processed (Art. 5 sect. 1 lit. b GDPR) or for which the AI system is used (see, in particular,

Art. 7 sect. 2 lit. a and Art. 8 sect. 2 AI Regulation draft); both laws require the regulation addresses to identify the risks to the fundamental rights based on the purpose of the data processing or AI system and to take technical and organisational measures to reduce these risks to a legally adequate level (Art. 1 sect. 2, 24, 25 and 32 GDPR; Art. 7 sect. 1 lit. b and Art. 9 sect. 2 AI Regulation draft); and if the purpose changes, this assessment starts all over again (Art. 6 sect. 4, as well as sect. 3 of Art. 13 and 14 GDPR; Art. 43 sect. 4 as well as Art. 28 sect. 1 lit. c AI Regulation draft). Of course, there are also differences. Three major differences are: Unlike the AI Regulation draft, the scope of application of the GDPR does not only depend on specific purposes or areas (defined in the law) and, thus, on specific risks to fundamental rights (that are predefined by the purpose in one of these areas), but rather unfolds its protection effect already in a precautionary stage (thus, before a certain purpose causes a specific risk to one or more fundamental rights) and through its cross-purpose and cross-sectoral approach (von Grafenstein 2021b). Second, the AI Regulation draft is primarily aimed at the providers (i.e. the developers) of AI systems and secondarily at the actual users of the systems as well as all intermediaries in between (Art. 16 et seq. AI Regulation draft). In this way, the legislator establishes liability of developers in the AI Regulation draft, which has long been demanded for the GDPR (Sydow-Mantz 2018). Finally, with the mutual information obligations of the provider, the users and intermediaries inform each other of any new risks that may arise (Art. 26 sct. 2, Art. 27 sect. 4, and 29 sect. 4 AI Regulation draft), the legislator seems to be taking up an approach that has also been applied by the REACH Regulation, according to which a major problem in the identification and control of risks is that the necessary information chain of all actors involved is not closed, at least not in a timely enough manner.[15] In the GDPR, this is only achieved with difficulty, especially through the now expanded construct of joint controllership, whose application in legal interpretation is fraught with many legal uncertainties (Gierschmann 2020).

While the GDPR focuses on the processing of data, the AI Regulation draft on the technology. This is complementary. However, both laws together run the risk of over-regulation.

In conclusion, this only slightly different scope and regulatory approach of the AI regulation draft compared to data protection might be disappointing. This may be especially true in view of the fact that there are several more areas of application in which personal data is not necessarily processed, apart from critical infrastructure, and therefore there may actually be a need for more protection: this may be especially true for risks to society as a whole, such as environmental protection, which cannot be adequately addressed at the individual fundamental rights level. Even worse, while both laws complement each other well in specific points, they also carry the obvious risk of largely redundant and thus disproportionate regulation (a problem that already arises for the legal system of the GDPR alone, von Grafenstein 2021c).

Against this background, it is all the more important to clarify the interplay of both laws especially where their compliance assessments overlap. But even in this respect, the reader is surprised. The AI regulation draft only contains two explicit provisions in this regard: Art. 10 sect. 5 allows the processing of even special categories of personal data if it is "strictly necessary for the purposes of ensuring bias monitoring, detection and correction in relation to the high-risk AI systems"; and Art. 29 sect. 6 requires the users of AI systems to use the information provided by the developer (Art. 13) to conduct an eventually necessary data protection impact assessment according to Art. 35 GDPR. Nevertheless, there is at least one other possibility (even if not explicitly mentioned) of further synergy effects in the compliance efforts, namely the synchronisation of the AI conformity assessments and, in particular, the certification mechanisms of the GDPR.

Delving deeper into this, it is first remarkable that the AI Regulation draft foresees two kinds of conformity

assessments: an internal self-assessment in the case that there are officially harmonised standards or common specifications of the AI Regulation requirements; and an external third-party assessment for the case that there are no (or not yet) such official standards or specifications (Art. 40-43 AI Regulation draft). In both cases, however, the provider as well as all other parties are forbidden to bring the AI system onto the market or use it unless the provider has gone through at least one of both conformity assessments (and has attached a CE marking to the system indicating the positive result of the assessment, Art. 19 sect. 1, Art. 26 sect. 1 lit. c, Art. 27 sect. 1 and 40 et seq.). In contrast, a GDPR-certification mechanism addresses processing operations and is, as mentioned before, voluntary. Thus, a controller (i.e. the primarily responsible entity) or processor (who processes the data on behalf of the controller) is free to get its processing operations certified. This means that both conformity assessments can complement each other not only because both laws address different actors: the AI Regulation draft addresses the developer of an AI system, while the GDPR addresses, in the terminology of the AI Regulation draft, the user. Rather, since the developer of an AI system is not allowed to place the system on the market without such an assessment, it is reasonable that a controller or processor who uses an AI system for its processing operations may base the certification of its processing operations on the conformity assessment of the AI system. The owner of the certification scheme or the certification body may therefore determine under which conditions and to what extent this can be done. However, it should be noted that conformity assessments, which do not meet the requirements of Art. 42 GDPR, can only be "taken into account" (instead of being directly applied). This means that the certification body can use such assessments as indications for compliance with the criteria of the present certification programme, but these do not replace the independent audit by the certification body.

In practice, the risk of redundant fundamental rights protection can be mitigated by GDPR certification schemes building on the AI systems conformity assessments.

On the other hand, there are also elements of the AI Regulation that should be placed on a broader footing in the entire system of the Digital Services Package. This refers to the system of mutual information obligations briefly outlined before, according to which the developer of an AI system, the user of this system and all intermediaries in between must inform each other about newly emerging risks in order to be able to react in time with the necessary countermeasures. Since, as shown, such unexpected risks can occur not only with AI systems or with the processing of personal data, but also with other risks (e.g. with respect to trade secrets, competitive disadvantages, the environment), such an information duty should be implemented across sectors and technologies. Therefore, data users should generally be obliged to systematically record factors of these risks in the respective data catalogues and potential countermeasures making it available to up and downstream data users (and data holders in the case of joint responsibilities). However, the right place for this obligation would be the Data Governance Act as it is a general requirement to maintain data quality and does not only refer to risks of an AI system or of processing personal data.

In contrast, a general obligation to record new risks (if of legal relevance) in data catalogues for downstream users should be implemented in the Data Governance Act across all sectors.

Last but not least, it should be pointed out that there seems to be no particular need, at least not at first sight, for synchronising the risk assessments in AI Regulation and the GDPR with the Digital Services Act. The Digital Services Act (DSA) addresses, amongst others, very large online platforms, which means "provider(s) of a hosting service which, at the request of a recipient of the service, stores and disseminates to the public information" (Art. 2 lit. h) and "which provide their services to a number of average monthly active recipients of the service in the Union equal to or higher than 45 million" (Art. 25). Art. 26 requires

very large online platforms to analyse, evaluate and mitigate the risks caused with respect to the following aspects: a) the dissemination of illegal content; b) any negative effects for the exercise of the fundamental rights; c) intentional manipulation of their service with an actual or foreseeable negative effect on the protection of public health, minors, civic discourse, or actual or foreseeable effects related to electoral processes and public security. Even though most of these risks are caused by the processing of personal data, the Digital Services Act addresses the *systemic* risks and not primarily the *individual* risks. In contrast, the GDPR addresses such systemic risks exclusively via individual fundamental rights (von Grafenstein 2021b). How primarily systemic risks are analysed and assessed is a very important question, which, however, must be dealt with in another contribution. In any case, the scope of application of both laws is complementary. In contrast, the interplay of the Digital Services Act with the AI Regulation already consists in the fact that the latter, even if the wording of the AI Regulation seems to address systemic risks as well (at least this is suggested by the title of the last group of cases "democratic processes"), it only covers individual risks, as the examples of the judiciary use cases suggest. At least at second glance, these "judiciary"use cases do not overlap with the use cases "civic discourse" or "electoral processes" listed in the DSA. A further investigation should therefore be omitted here.

3.2.5 A claim for more solution-oriented regulation based on evidence

In the overall view of all these draft laws, it is first noticeable that the legislator solves the value-for-risk dilemma in data sharing in most sharing constellations with data access rights. In principle, this includes the sharing of data from public to private bodies, from private to public bodies and between private bodies insofar as it concerns data that has arisen through the use of a product or related service. Some laws seem a little weak in this respect. For example, the open data regulations for public bodies should be supplemented by an obligation to catalogue their data and a right of access for private bodies under the condition that all protection laws are met. In any case, since the legislator still requires compliance with conflicting protection laws, but does not itself specify the concrete technical and organisational measures required for their solution, this will lead to a plethora of disputes that will likely be fought out before the dispute settlement bodies, competent authorities and courts. In the best case, these bodies and authorities systematically prepare these cases with regard to their typified conflicts of interest and solutions across the whole EU Single Market – and publish them (see for data governance structures discussed with respect to the interplay of structures between private parties and the competent authorities at Prufer and Graef 2021). The proposed framework may provide analytical support in this regard. However, to proactively analyse, develop and propose concrete solutions, these competent bodies would need to be resourced accordingly. Otherwise, most access rights will come to nothing in practice.

Not least in the other data sharing constellations, the legislator relies on soft law measures. In this context, all possible data intermediation services – be it selectively on single data governance layers or combining legal, organisational and technological aspects – may help at least with the control and minimisation of compliance risks. Particularly with regard to so-called data sharing services, however, the legislator should reflect the concrete needs of the stakeholders involved in data sharing more precisely and implement them more stringently in the form of appropriate incentive and protection mechanisms. Above all, it should not unnecessarily reduce the remaining scope of sharing services for data-driven innovation by an excessive ban on using the mediated data for own purposes, but rather for the data holder's and data user's low-risk purposes.

In addition, legislators should put more effort into thinking through the interplay of, especially, overlapping protection laws. The extensive overlap between the AI Regulation draft and the GDPR is

perplexing and disappointing at the same time. Societal threats posed by AI systems are not adequately addressed by a regulatory approach that, like the GDPR, focuses on the level of individual fundamental rights. At the same time, data protection law gets an important supplement by the provider liability now provided for in the AI Regulation draft. However, the vast overlap of both protection systems overall also threatens to be redundant and disproportionate. Where the scopes of protection overlap, a systematic analysis and streamlining of the variety of both laws' protection instruments would be very much in need. To reach this aim, the synchronisation of the conformity assessment from the AI Regulation draft and GDPR certification procedures can only be a start. On the other hand, data users should generally be obliged to record factors of (all kinds of legally relevant) risks in their data catalogues and thus systematically collect knowledge about the potential risk of data use and make it available to up and downstream users (and data holders if shared responsibilities demand for this). However, the right place for this obligation would be the Data Governance Act as it is a general requirement to maintain data quality and does not only refer to risks of an AI system or of processing personal data.

3.3 So much for the regulator, what about the other actors?

In the very end, the framework presented here may make an analytical contribution to determining what each actor involved in the collection, reuse and/or sharing of data can and should do concretely across all data governance dimensions in order to fully unleash the potential of data-driven innovation. Ideally, the data governance framework helps to improve the empirical basis for more effective, evidence-based regulation. On this basis, this paper has made an initial overview analysis of some of the legislative proposals from the digital services package. However, it is not only the state regulator that has to play its supporting role in the success of data governance. First and foremost, it is the private and public bodies who need to recognise their mutual interests and resulting conflicts for successful data governance and the interdependencies for their solution. To do this, they need to find a common terminology, align their respective metrics, methods and procedures, and clarify their responsibilities. The place where these coordination efforts accumulate is probably the data catalogues, where the actors involved have to add their constantly generated knowledge about the value and risks of the data and the necessary technical-organisational measures. This may be tedious but unavoidable, at least if one wants to achieve the potential of data-driven innovation from the perspective of all stakeholders concerned by data-driven innovation, not only from a couple of them or even only one. On closer inspection, however, stakeholders can actually use the dependencies as a competitive advantage. This is especially true when their buyers, contractors, etc. are held legally (or at least politically) accountable to meet certain standards, but technically only their vendors or providers are able to do so. Then these sellers or providers can take advantage of this circumstance and offer such products or services that can be used according to those standards. In this case, the noble claim of the legislator to promote innovation and even create competitive advantages through regulation would actually get a step closer to reality. However, this requires an appropriate data strategy.

ENDNOTES

1) See, for instance, in Art. 2 no. 8 of the draft of the Data Governance Act (European Commission 2020b): "'access' means processing by a data user of data that has been provided by a data holder, in accordance with specific technical, legal or organisational requirements, without necessarily

implying the transmission or downloading of such data"; or Art. 11 no. 7: "the provider shall put in place adequate technical, legal and organisational measures in order to prevent transfer or access to non-personal data that is unlawful under Union law"; or Art. 30 sect. 1: "The public sector body, the natural or legal person (...) shall take all reasonable technical, legal and organisational measures in order to prevent transfer or access to non-personal data held in the Union (...)."

2) Examples of this include the extensive development of copyright law and its ancillary rights, such as database protection and the ancillary right for publishers, the General Data Protection Regulation (2016), as well as the Digital Services Package with its Data Governance Act (2020b), the Digital Markets Act (2020d), the Digital Services Act (2020c), AI Regulation, and the Data Act that is already in the legislative pipeline.

3) See, for example, the discussion on the so-called risk-based approach in data protection law, where the subjectivity of risk-perceptions has been solved by recurring to the legal concept of "risks to fundamental rights", since the fundamental rights can serve as an objective scale for the risk assessment (see in detail von Grafenstein 2021b, pp. 190-205).

4) See Art. 4 no. 8 GDPR (2016).

5) See, for instance, https://new.siemens.com/global/de/produkte/services/iot/city-performance-tool.html.

6) However, see the criticism of disproportionate regulatory effects of the GDPR (von Grafenstein 2021c)

7) See https://www.oracle.com/big-data/what-is-a-data-catalog/.

8) See this observation, still relatively undifferentiated though, at von Grafenstein et al. (2019); focusing on the challenge of open-ended innovation processes (von Grafenstein 2018); cf. similar reasons mentioned by V Kathuria (2019).

9) From the perspective of information science, for example, Becker et al. (2021), as well as from a risk regulation perspective von Grafenstein (2020b).

10) Concrete but veiled risks exist, for example, in the case of personalised advertising for privacy, autonomous consumer goods purchasing decisions and freedom from discrimination, see von Grafenstein (2018, pp. 624 et seq.)

11) In the end, this is similar to Ladley (2019).

12) See in favour of decentralised structures, in general, Gagzow (2018, pp. 61 et seq.), with further references, online available at: https://www.datenschutzzentrum.de/uploads/projekte/ikopa/iKoPA_D3.2-3.pdf; see in more detail in the research project "How to Build Data-Driven Innovation Projects at Large with Data Protection by Design" (von Grafenstein 2020c), where the research group proposes a "fully distributed" architecture, online accessible under: https://www.hiig.de/publication/paper-data-protection-by-design-in-smart-cities/.

13) Such a transparency obligation for personal data is already stipulated by the GDPR, as are all the other data protection measures that a data intermediary has to apply.

14) Of course, the importance of the trust-building function of data intermediaries requires such certification schemes to be sufficiently specific to avoid potential loopholes for misuse, see von Grafenstein (2021a, pp. 4 and 16).

15) See, for instance, https://ec.europa.eu/environment/chemicals/reach/reach_en.htm#:~:text=REACH%20(EC%201907%2F2006),authorisation%20and%20restriction%20of%20chemicals

REFERENCES

Abbasi A, Sarker S, and Chiang R (2016) Big Data Research in Information Systems: Toward an Inclusive Research Agenda. Journal of the Association for Information Systems 17(2) doi: 10.17705/1jais.00423

Abraham R, Schneider J, vom Brocke J (2019) Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management* 49, 424–438 doi: 10.1016/j.ijinfomgt.2019.07.008

Acquisti, A, Brandimarte, L, and Loewenstein, G (2020). Secrets and Likes: The Drive for Privacy and the Difficulty of Achieving It in the Digital Age. Journal of Consumer Psychology 30(4), pp. 736–758.

Albers M (2012) § 22 – Umgang mit personenbezogenen Informationen und Daten. In: Hoffmann-Riem W, Schmidt-Aßmann E, Voßkuhle A (eds) *Grundlagen des Verwaltungsrechts. Band II: Informationsordnung - Verwaltungsverfahren - Handlungsformen*, 2nd edn. C.H. Beck, 107-234, München

Art. 29 Data Protection Working Party (2013) *Opinion 03/2013 on purpose limitation*

Baldwin R, Cave M, Lodge M (2013) *Understanding regulation: theory, strategy, and practice* , 2nd edn. Oxford University Press on Demand, Oxford

Blankertz A and & Specht L (2021) Policy Brief: What regulation for data trusts should look like. Stiftung Neue Verantwortung: https://www.stiftung-nv.de/en/publication/regulation-data-trusts

Becker C, Antunes G, Barateiro J, Vieira R, Borbinha J (2012) Control objectives for DP: Digital preservation as an integrated part of IT governance. *American Society for Information Science and Technology* 1(48), 1-10 doi: 10.1002/meet.2011.14504801124

Beierle TC (2002) The Quality of Stakeholder-Based Decisions. *Risk Analysis* 22(4), 739–749 doi: 10.1111/0272-4332.00065

Bélanger F, James T (2020) A Theory of Multilevel Information Privacy Management for the Digital Era. Information Systems Research 31(2), pp. 510-536 doi: 10.1287/isre.2019.0900

Beynon-Davies P, Wang Y (2019) Deconstructing Information Sharing. *Journal of the Association for Information Systems* 20(4) doi: 10.17705/1.jais.00541

Bitkom (2020) *Anonymisierung und Pseudonymisierung von Daten für Projekte des maschinellen Lernens - Eine Handreichung für Unternehmen.* https://www.bitkom.org/sites/main/files/2020-10/201002_lf_anonymisierung-und-pseudonymisierung-von-daten.pdf

Black J (2014) Learning from Regulatory Disasters. *Policy Quarterly 3(10)* doi: 10.26686/pq.v10i3.4504

Britz G (2010) Informationelle Selbstbestimmung zwischen rechtswissenschaftlicher Grundsatzkritik und Beharren des Bundesverfassungsgerichts. *Offene Rechtswissenschaft*, 561–596

Braithwaite J, Coglianese C, Levi-Faur D (2007) Can regulation and governance make a difference? *Regulation & Governance* 1(1), 1-7 doi: 10.1111/j.1748-5991.2007.00006.x

Brous P, Janssen M (2020) Trusted Decision-Making: Data Governance for Creating Trust in Data Science Decision Outcomes. *Administrative Sciences* 10(4):81 doi: 10.3390/admsci10040081

Bundesministerium für Bildung und Forschung (2021b) *Förderung von Datentreuhandmodellen - BMBF Digitale Zukunft*

Bygrave LA (2020) Article 25. Data protection by design and by default. In: Kuner C, Bygrave LA, Docksey C (eds) *The EU General Data Protection Regulation (GDPR): A Commentary*, 571-581. Oxford University Press

Chesbrough HW (2003) *Open innovation: The new imperative for creating and profiting from technology*. Harvard Business Press

Contreras JL, Reichman JH (2015) Sharing by design: Data and decentralized commons. *Science* 350(6266), 1312–1314 doi: 10.1126/science.aaa7485

Die Bundesregierung (2021) *Datenstrategie der Bundesregierung*

Docksey C (2020) Article 24. Responsibility of the controller. In: Kuner C, Bygrave LA, Docksey C (eds) *The EU General Data Protection Regulation (GDPR): A Commentary,* 554-570. Oxford University Press

Elliot M, Mackey E, O'Hara K, Tudor C (2016) *The Anonymization Decision-Making Framework*. United Kingdom Anonymisation Network. http://eprints.soton.ac.uk/id/eprint/399692 Accessed 12 Oct 2021

European Commission (2020a) *Communication of the EU Commission, A European Data Strategy.* Brussels

European Council (2022), Digital Markets Act (DMA): agreement between the Council and the European Parliament. Press release 25 March 2022 00:05

Foster J, McLeod J, Nolin J, Greifeneder E (2018) Data work in context: Value, risks, and governance. *Journal of the Association for Information Science and Technology* 69(12), 1414–1427 doi: 10.1002/asi.24105

Frank R D, von Grafenstein M & Rothfritz L (2022). Open Data und die Risikowahrnehmung in der Öffentlichen Daseinsvorsorge. Zenodo. DOI: 10.5281/zenodo.6285549

Frey C B, Presidente G (2022) The GDPR effect: How data privacy regulation shaped firm performance globally. VOX EU.
https://voxeu.org/article/how-data-privacy-regulation-shaped-firm-performance-globally

Friederici N, Krell T, Meier P, Braesemann F, Stephany F (2019) *Plattforminnovation im Mittelstand Abschlussbericht des Forschungsvorhabens* fe 12/19: „Hindernisse und Gelingensbedingungen für kooperative Ansätze kleiner und mittlerer Unternehmen in datenbasierten Märkten und Branchen"

Friedman AL, Miles S (2006) *Stakeholders: Theory and Practice*. Oxford University Press

Gagzow G (2018) Integrierte Kommunikationsplattform für automatisierte Elektrofahrzeuge - Deliverable 3.2 Datenschutz bei vernetzten, automatisierten und kooperativen Fahrzeugen nach der Datenschutzgrundverordnung, Version 1.0, Unabhängiges Landeszentrum für Datenschutz

Gangadharan S (2014) Data and discrimination: Collected essays. Washington DC: Open Technology Institute

Gierschmann S (2020) Gemeinsame Verantwortlichkeit in der Praxis. Systematische Vorgehensweise zur Bewertung und Festlegung. *Zeitschrift für Datenschutz* (2), 69.

Graef I, Gellert R (2021) The European Commission's proposed Data Governance Act: some initial reflections on the

increasingly complex EU regulatory puzzle of stimulating data sharing. TILEC Discussion Paper No. DP2021-006, Available at SSRN: https://ssrn.com/abstract=3814721 or http://dx.doi.org/10.2139/ssrn.3814721

Günther WA, Rezazade Mehrizi MH, Huysman M, Feldberg F (2017) Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems* 26(3), 191–209 doi: 10.1016/j.jsis.2017.07.003

Habermas J (1995) *Theorie des kommunikativen Handelns*, Band 2: Zur Kritik der funktionalistischen Vernunft, 10. Suhrkamp

Hansen M (2019a) DSGVO Art. 25 Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellungen. In: Simitis, Hornung, Spiecker gen. Döhmann (eds) *Datenschutzrecht*, 746-766. 1st edn. Nomos Verlag, Baden-Baden

Hansen M (2019b) DSGVO Art. 32 Sicherheit der Verarbeitung. In: Simitis, Hornung, Spiecker gen. Döhmann (eds) *Datenschutzrecht*, 813-835. 1st edn. Nomos Verlag, Baden-Baden

Hassan N, Rivard S, Lowry P, Matthiassen L (2022) Useful Products in Information Systems Theorizing: A Discursive Formation Perspective. Journal of the Association for Information Systems doi: 10.17705/1jais.00730

Hofmann J, Katzenbach C, Gollatz K (2017) Between coordination and regulation: Finding the governance in Internet governance. *New Media & Society* 19(9), 1406–1423

Hölzel J (2019) Differential Privacy and the GDPR. *European Data Protection Law Review* 5(2), 184–196 doi: 10.21552/edpl/2019/2/8

Janssen M, Brous P, Estevez E, Barbosa LS, Janowski T (2020) Data governance: Organizing data for trustworthy Artificial Intelligence. *Government Information Quarterly* 37(3):101493 doi: 10.1016/j.giq.2020.101493

Kahanwal DB, Singh DTP (2012) The Distributed Computing Paradigms: P2P, Grid, Cluster, Cloud, and Jungle. *International Journal of Latest Research in Science and Technology* 1(2), 183–187

Kathuria V (2019) Greed for data and exclusionary conduct in data-driven markets. *Computer Law & Security Review* 1(35), 89-102 doi: 10.1016/j.clsr.2018.12.001

Khatri V, Brown CV (2010) Designing data governance. *Communications of the Association for Computing Machinery* 53(1), 148–152 doi: 10.1145/1629175.1629210

Konferenz der unabhängigen Datenschutzbehörden des Bundes und der Länder (2019) *Das Standard-Datenschutzmodell – Eine Methode zur Datenschutzberatung und -prüfung auf der Basis einheitlicher Gewährleistungsziele,* Version 2.0a

Krotova A, Eppelsheimer J (2019) Was bedeutet Data Governance? Eine Clusteranalyse der wissenschaftlichen Literatur zu Data Governance, *Demand Project - Data Economics and Management of Data Driven Business*

Krotova A, Spiekermann M (2020) Data Valuation Model: Handbuch für Bewertung von Daten in Unternehmen , *Demand Project - Data Economics and Management of Data Driven Business*

Ladley J (2019) *Data governance: How to design, deploy, and sustain an effective data governance program*. Academic Press

Loshin D (2008) Master Data Management. Elsevier Science & Technology, San Francisco

Matthijs M, Parsons C, Springer B (2021) *Why Did Europe's Single Market Surpass America's?* Presented at the 10th Conference of the ECPR Standing Group on the European Union, June 10-12, 2021

Mayer-Schönberger V, Cukier K (2013) *Big data: A revolution that will transform how we live, work, and think* . Houghton Mifflin Harcourt

Mousavi S, Gigerenzer G (2014) Risk, uncertainty, and heuristics. *Journal of Business Research* 67(8), 1671–1678 doi:

10.1016/j.jbusres.2014.013

Müller G, Flender C, Peters M (2012) Vertrauensinfrastruktur und Privatheit als Ökonomische Fragestellung. In: Buchmann J (ed) *Internet Privacy: Eine multidisziplinäre Bestandsaufnahme/A multidisciplinary analysis*, 143–188. Springer, Berlin, Heidelberg

Osterwalder A, Pigneur Y (2010) *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers.* John Wiley & Sons

Otto B (2011a) Data Governance. *Wirtschaftsinformatik* 53(4), 235–238 doi: 10.1007/s11576-011-0275-1

Otto B (2011b) Organizing Data Governance: Findings from the Telecommunications Industry and Consequences for Large Service Providers. *Communications of the Association for Information Systems* 29(1) doi: 10.17705/1CAIS.02903

Petri T (2019) DSGVO Art. 24 Verantwortung des für die Verarbeitung Verantwortlichen. In: Simitis, Hornung, Spiecker gen. Döhmann (eds) *Datenschutzrecht*, 739-746. 1st edn. Nomos Verlag, Baden-Baden

Pohle J (2018) D*atenschutz und Technikgestaltung: Geschichte und Theorie des Datenschutzes aus informatischer Sicht und Folgerungen für die Technikgestaltung*. PhD diss, Mathematisch-Naturwissenschaftliche Fakultät, Humboldt-Universität zu Berlin

Pombriant D (2013) Data, Information and Knowledge: Transformation of data is key. *Computer law review international* (4), 97–102 doi: 10.9785/ovc-cri-2013-97

Prufer, J and Graef, I, Governance of Data Sharing: a Law & Economics Proposal (January 22, 2021). TILEC Discussion Paper No. 2021-001, CentER Discussion Paper No. 2021-004, Available at SSRN: https://ssrn.com/abstract=3774912 or http://dx.doi.org/10.2139/ssrn.3774912

Riemensperger F and Falk S (2019) *Titelverteidiger: Wie die deutsche Industrie ihre Spitzenposition auch im digitalen Zeitalter sichert*. Münchner Verlagsgruppe GmbH, München.

Sarkar S, Banatre J-P, Rilling L, Morin C (2018) Towards Enforcement of the EU GDPR: Enabling Data Erasure. *Institute of Electrical and Electronics Engineers*, 222–229 doi: 10.1109/Cybermatics_2018.2018.00067

Smallwood RF (2014) Information Governance Principles. In: *Information Governance: Concepts, Strategies, and Best Practices*, 25–41. John Wiley & Sons, Hoboken, New Jersey

Sorescu A (2017) Data-Driven Business Model Innovation. *Journal of Product Innovation Management* 34(5), 691–696 doi: 10.1111/jpim.12398

Spiekermann M (2019) Data Marketplaces: Trends and Monetisation of Data Goods. *Intereconomics* 54(4), 208–216 doi: 10.1007/s10272-019-0826-z

Sydow G (2018) *Europäische Datenschutzgrundverordnung: Handkommentar.* 2. Auflage, Nomos Verlag

Tallon P, Ramirez R, Short J (2013) The Information Artifact in IT Governance: Toward a Theory of Information Governance. *Journal of Management Information Systems* (30), 141–178 doi: 10.2753/MIS0742-1222300306

van de Ven M, Abbas A, Kwee Z, Reuver M (2021) *Creating a Taxonomy of Business Models for Data Marketplaces*, 34th Bled eConference: Digital Support from Crisis to Progressive Change - Online, Bled, Slovenia

Vollmer N (2021) *Recital 78 EU General Data Protection Regulation (EU-GDPR).* https://www.privacy-regulation.eu/en/recital-78-GDPR.htm. Accessed 10 Oct 2021

von Grafenstein M (2020a) The EU General Data Protection Regulation Outside the Box: Competitive Advantages and Openness to Innovation. In: Jakobi et al. (2020) *The Role of IS in the Conflicting Interests Regarding GDPR*. *Business & Information Systems Engineering* 62(3), 261–272 doi: 10.1007/s12599-020-00633-4

von Grafenstein M (2020b) Refining the Concept of the Right to Data Protection in Article 8 ECFR - Part I: Finding an Appropriate Object and Concept of Protection by Re-Connecting Data Protection Law with the Concepts of Risk Regulation. European Data Protection Law Review 04/2020 (Vol.6), 509-521 doi: 10.21552/edpl/2020/4/7

von Grafenstein M (2021b) Refining the Concept of the Right to Data Protection in Article 8 ECFR – Part II: Controlling Risks through (Not To) Article 8 ECFR against the Other Fundamental Rights (Esp. by the Principle of Purpose Limitation) *European Data Protection Law Review* 02/2021 (Vol. 7), 190-205 doi: 10.21552/edpl/2021/2/8

von Grafenstein M (2021c) Refining the Concept of the Right to Data Protection in Article 8 ECFR - Part III: Consequences for the interpretation of the GDPR (and the lawmaker's room for maneuver). *European Data Protection Law Review* 03/2021 (Vol. 7)

von Grafenstein M (2021d). Specific certification schemes as rule, general schemes (and criteria) as exception. HIIG Discussion Paper Series, 2021(04). DOI: 10.5281/zenodo.4905484

von Grafenstein M (2022) Co-Regulation and the Competitive Advantage in the GDPR: Data Protection Certification Mechanisms, Codes of Conduct and the 'State of the Art' of Data Protection-by-Design. Forthcoming in Gonzàles-Fuster G, van Brakel R, De Hert P (eds.) *Research Handbook on Privacy and Data Protection Law. Values, Norms and Global Politics*, Edward Elgar Publishing. Available at SSRN: https://ssrn.com/abstract=3336990

von Grafenstein M, Jakobi T, Stevens G (in review) Effective Data Protection by Design through interdisciplinary research methods - The example of effective purpose specification by applying user-centered UX-design methods. *Computer Law and Security Review*

Wernick A, Olk C, von Grafenstein M (2020) Defining Data Intermediaries: A Clearer View Through the Lens of Intellectual Property Governance. *Technology and Regulation,* 65–77 doi: 10.26116/techreg

Ye Q, Hu H (2020) Local differential privacy : tools, challenges, and opportunities. In: *Communications in computer and information science,* 13-23 doi: 10.1007/978-981-15-3281-8_2

Yeung K (2017) Algorithmic Regulation: A Critical Interrogation. Regulation & Governance 12(4), 505-523 doi: 10.1111/rego.12158

Gaia-X: A Federated Secure Data Infrastructure. *https://www.gaia-x.eu/* Accessed 10 Oct 2021

LAWS CITED

Data Governance Act (2020b) COM/2020/767 final

Digital Services Act (2020c) COM/2020/825 final

Digital Markets Act (2020d) COM/2020/842 final

Data Act (2022) COM/2022/68 final

Freedom of Access to Information (2003) OJ L 41, 14.2.2003, p. 26-32

General Data Protection Regulation (2016b) OJ L 119, 4.5.2016,. 1-88

Trade Secret Directive (2016a) OJ L 157, 15.6.2016, p. 1-18