COMPUTER LAW & SECURITY REVIEW 46 (2022) 105722

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/CLSR

Computer Law & Security Review

Effective data protection by design through interdisciplinary research methods: The example of *effective* purpose specification by applying user-Centred UX-design methods ☆,☆☆

Max von Grafenstein[a,b,1,*], *Timo Jakobi*[c,d,e,2], *Gunnar Stevens*[f,g,3]

[a] *ECDF / Berlin University of the Arts, Bundesallee 1-12, 10719, Berlin, Germany*
[b] *HIIG, Französische Str. 9, 10117, Berlin, Germany*
[c] *ECDF / Berlin University of the Arts, Adresse, Germany*
[d] *University of Siegen, Kohlbettstraße 15, 57072, Siegen, Germany*
[e] *University of Applied Sciences Bonn-Rhein-Sieg, Grantham-Allee 15, 53757, Sankt Augustin, Germany*
[f] *University of Siegen, Kohlbettstraße 15, 57072, Siegen, Germany*
[g] *University of Applied Sciences Bonn-Rhein-Sieg, Grantham-Allee 15, 53757, Sankt Augustin, Germany*

ARTICLE INFO

Keywords:
Data protection by design
Effective purpose specification
GDPR
UXD
HCI

ABSTRACT

While the recent discussion on Art. 25 GDPR often considers the approach of data protection by design as an innovative idea, the notion of making data protection law more effective through requiring the data controller to implement the legal norms into the processing design is almost as old as the data protection debate. However, there is another, more recent shift in establishing the data protection by design approach through law, which is not yet understood to its fullest extent in the debate. Art. 25 GDPR requires the controller to not only implement the legal norms into the processing design but to do so *in an effective manner*. By explicitly declaring the effectiveness of the protection measures to be the legally required result, the legislator inevitably raises the question of which methods can be used to test and assure such efficacy. In our opinion, extending the legal compatibility assessment to the real effects of the required measures opens this approach to interdisciplinary methodologies. In this paper, we first summarise the current state of research on the methodology established in Art. 25 sect. 1 GDPR, and pinpoint some of the challenges

☆ The example of *effective* purpose specification by applying user-Centred UX-design methods
☆☆ Referee suggestions: Lee Bygrave – XXX
* Corresponding author at: Jungstr. 29, 10247 Berlin, Germany.
E-mail addresses: m.von-grafenstein@udk-berlin.de, max.grafenstein@hiig.de (M. von Grafenstein).
[1] equal lead author, legal background
[2] equal lead author, Human Computer Interaction background
[3] third author, Prof. for Consumer Informatics

https://doi.org/10.1016/j.clsr.2022.105722

of incorporating interdisciplinary research methodologies. On this premise, we present an empirical research methodology and first findings which offer one approach to answering the question on how to specify processing purposes effectively. Lastly, we discuss the implications of these findings for the legal interpretation of Art. 25 GDPR and related provisions, especially with respect to a more effective implementation of transparency and consent, and provide an outlook on possible next research steps.

1. Introduction: how to assess effective measures of data protection by design?

The concept of Data Protection by Design, as provided for in Article 25 sect. 1 of the EU General Data Protection Regulation (GDPR), is seen as an innovative regulatory approach. In short, the concept obliges data controllers to implement all legal requirements of the GDPR into the technical and organisational design of their data processing so that the rights and freedoms of data subjects are effectively protected against the processing risks. In fact, the idea of making data protection law more effective by requiring the regulation addressee to implement the legal norms into the technical and organisational design of the data processing activities is rather old. Actually, this concept is nearly as old as the data protection and privacy debate. However, there is a more recent shift towards establishing the data protection by design approach within law, and the implications of such an approach are not yet fully understood in the data protection debate. As mandated by Article 25 GDPR, the regulation addressee is required not only to implement the legal norms into the processing design, but to do so in an effective manner. By explicitly declaring the effectiveness of the protection measures as a legally required result, the legislator inevitably raises the question of which methods can be used to test and ensure effectiveness. In fact, extending the legal conformity assessment to the real effects of the required measures opens this assessment to (non-legal) methodologies which are specialised for assessing data protection measures empirically. This shift towards using non-legal methodologies for the purposes of reaching legal compliance marks a fundamental paradigm change in the application of the law. On one hand, the introduction of interdisciplinary measures may lead to an increase in complexity; on the other hand, it may also positively contribute to a rationalisation of the law. In this paper, we take the principle of purpose limitation to demonstrate how this paradigm shift, as enshrined in Art. 25 GDPR, can be implemented practically. The contribution of this paper therefore is twofold: first, we argue in favour of using empirical methods to assess the effectiveness of data protection measures. Second, we demonstrate how to test and assure the effective formulation of processing purposes and how improved processing purpose specifications may look like in future everyday practice. Effective purpose specifications in particular, but also the basic method presented here, are a prerequisite for the effective implementation of numerous other data protection requirements, above all effective transparency and consent mechanisms. These include, not least, privacy icons, which are intended by the legislator to contribute significantly to the comprehensibility of these measures (see Art. 12 sect. 7 GDPR).

1.1. *The idea and history of data protection by design*

Using the data protection by design approach to effectively specify processing purposes might be surprising. Indeed, many authors assess this approach through the prism of technical and organizational measures (eg. pseudonymization) and consider the processing operation itself untouchable. Such concerns are the result of misunderstandings of the data protection by design approach, and these misperceptions are almost as old as the idea of data protection by design.

Back in 1969, Miller had already claimed that technical safeguards "will be most efficient and economical if they are incorporated into the original design of the hardware and software than if they are added subsequently".[4] However, while Miller's work focused on data security, his argument can likewise be applied to data protection.[5] Based on the ensuing discussion, the German Bundestag was the first parliamentary body worldwide to enact a provision incorporating a data protection by design approach. The resulting Federal Data Protection Law of 1978 provided a provision (i.e. § 6) which shared certain similarities to Art. 25 GDPR:

> "*Whoever processes personal data within the scope of Section 1 (2) or on behalf of the persons or bodies mentioned therein shall take the technical and organisational measures necessary to ensure the implementation of the provisions of this Act, in particular the requirements specified in the Annex to this Act. Measures shall only be necessary if their cost is in reasonable proportion to the regulatory goal.*"[6]

[4] A. R. Miller, 'Personal Privacy in the Computer Age: The Challenge of a New Technology in an Information-Oriented Society' (1969) 67 Michigan Law Review, p. 1089.

[5] J. Pohle, 'Das Scheitern von Datenschutz by Design: Eine kurze Geschichte des Versagens' (2015) 2 FlfF-Kommunikation 32, pp. 41-44.

[6] Quoted according to J. Pohle, 'Das Scheitern von Datenschutz by Design: Eine kurze Geschichte des Versagens' (2015) 2 FlfF-Kommunikation 32, p. 42: „Wer im Rahmen des § 1 Abs. 2 oder im Auftrag der dort genannten Personen oder Stellen personenbezogene Daten verarbeitet, hat die technischen und organisatorischen Maßnahmen zu treffen, die erforderlich sind, um die Ausführung der Vorschriften dieses Gesetzes, insbesondere die in der Anlage zu diesem Gesetz genannten Anforderungen zu gewährleisten. Erforderlich sind Maßnahmen nur, wenn ihr Aufwand in einem angemessenen Verhältnis zu dem angestrebten Schutzzweck steht."

Surprisingly, the legal scope of this provision was quickly reduced to an issue of data security. German data protection authorities concluded that § 6 only required data security measures "which aim at a trouble-free data processing secured against misuse".[7] The data protection theorist Pohle rightly considered this interpretation by authorities as an abandonment of the original legislative goal: Data protection by design is not limited to the technological level of data security, but refers to the design of the entire processing operation. This includes, but is not limited to, the organisational and technological design within which the processing operation is embedded.[8] To understand this difference to its fullest extent, one must consider the conceptual difference between data protection and data security. Data security focuses on, among other aspects, the protection of confidentiality against unauthorized access (which often comes from an outside party to the processing entity) to information that is considered "private" for whatever reason.[9] In contrast, data protection addresses the informational power asymmetries that arise through the use of information technologies (usually from within the processing entity). The more widespread information technologies become these days, for example, the easier may an employer collect, combine and process information about its (potential) employees to evaluate their suitability for certain jobs or their performances without giving them the chance to know what is going on, to verify the accuracy of the data, to be able to counter the conclusion, or otherwise react to it in any way. These informational power asymmetries are not intrinsically illegitimate, but can result in a use of gathered information that undermines the data subject's rights to freedom and participation. In this concept, the specification of the processing purpose plays a central role because, due to the IT-enabled variety of possible data uses, there is no longer data that is irrelevant per se; rather, the relevance of data is determined largely by its processing purpose.[10] Data protection reacts to this observation by requiring the controller to specify its processing purpose and, on this basis, implement measures which accordingly limit the use of the data, make it transparent to the data subject and enable them to intervene if they have objections.[11] Correspondingly, empirical research shows how transparency and control are driving factors which facilitate user trust and ultimately technology acceptance in and beyond systems that process personal data.[12] Thus, technological data security measures, which may exclude an unauthorized access to the data, support the goal of technology acceptance, but are not sufficient to control (or mitigate) the aforementioned threats of informational power asymmetries. Consequently, the data protection by design approach includes, with its holistic view of the personal data processing embedded in its technical-organisational system, all components that increase (or decrease) the informational power asymmetry.

The background context we have provided explains the original objective of the data protection by design approach and, consequently, why the approach established in Art. 25 GDPR cannot be reduced to technical-organisational measures (like pseudonymisation or encryption).[13] This is also why Art. 25 GDPR requires the controller to implement *all* data protection principles under Art. 5 GDPR in its processing design. Thus, Art. 25 GDPR does not only include the principle of confidentiality (e.g. requiring technical encryption), but also purpose limitation (including purpose specification) as well as all other principles and further legal requirements, which mostly depend, in fact, on how the processing purpose is specified: for example, the purpose specified by the controller determines the appropriate legal basis (lawfulness principle), what data are necessary and how long the data must be stored (principle of data minimisation and storage limitation), what the controller has to make transparent (transparency principle) and so on; likewise, the specification of the purpose determines the data subject's intervention rights including their rights to consent, object, data access, correction, and deletion. To the extent that these measures require a human computer interface, the design of this interface must also comply with the data protection by design-approach. In any case, the effective implementation of most of these GDPR provisions presupposes an effective specification of the purposes. For this reason, the question of effective purpose specification is the focus of this contribution. As such, this contribution can provide

[7] J. Pohle, 'Das Scheitern von Datenschutz by Design: Eine kurze Geschichte des Versagens' (2015) 2 FlfF-Kommunikation 32, p. 42, quoting the version of the provisions in the Official Gazzette of the Land Hessen, Nr. 50, 11.12.1978, pp. 2451-2457: "Damit seien Maßnahmen gemeint, 'die eine störungsfreie und gegen Mißbrauch gesicherte Datenverarbeitung zum Ziel haben' (I.1.1)".

[8] J. Pohle, 'Das Scheitern von Datenschutz by Design: Eine kurze Geschichte des Versagens' (2015) 2 FlfF-Kommunikation 32, p. 43.

[9] Cf. European Court of Human Rights, 'ECHR 2008/14 Case of I v. Finland, 17 July 2008, No. 20511/03 (Fourth Section)' (2008) 15 European Journal of Health Law 426, p. 47.

[10] In the German legal-conceptual development of data protection during the 70ies and 80ies, this focus on the purpose represented an explicit departure from the classic privacy concept, which until then had been based primarily on the context in which the data was collected (e.g. in the public, private or intimate context), see M. Albers, Informationelle Selbstbestimmung, Baden-Baden: Nomos, 2005, pp. 211 and 212; this is not to say that the debate that primarily in the USA has already started in the 60ies and continues to be conducted under the privacy label, has not led to similar developments, which nowadays equally take into account the context in which the data is *used*, see for example H. Nissenbaum, Privacy in Context, Technology, Policy, and the Integrity of Social Life, Stanford University Press 2009.

[11] M. von Grafenstein, '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)', going to be published in EDPL 01/2021, with further references.

[12] P. Pavlou,' Integrating trust in electronic commerce with the technology acceptance model: model development and validation' (2001) 159 AMCIS 2001; Transparency, Privacy and Trust – Technology for Tracking and Controlling My Data Disclosures: Does This Work?

[13] Cf. EDPB, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (adopted on 20 October 2020), version 2.0, cip. 7 et seqq.; EDPS, 'European Data Protection Supervisor, Opinion 7/2015, Meeting the Challenges of Big Data, A Call for Transparency, User Control, Data Protection by Design and Accountability' (2015), p. 14; in contrast, denying the application of an holistic approach, J. Moser 'Kommentar Datenschutz-Grundverordnung' (2017, 1st edn. Bundesanzeiger Verlag), cip. 33.

an important basis for subsequent research on the effective implementation of further GDPR provisions, such as more effective transparency and consent-mechanisms (including the currently much discussed privacy icons).[14] Against this background, it remains an open question as to why some data protection authorities still maintain a primarily security-focused interpretation of data protection and the risks data protection seeks to address.[15] As we have argued, this narrow interpretation loses sight of the main objective of data protection: the prevention of the potential abuse of informational power asymmetries, for which effective purpose specification plays a fundamental role.

1.2. *Clarifying the normative structure of Art. 25 sect. 1 GDPR*

Thus, data protection by design is not new under the GDPR, but rather the outcome of an ongoing debate for over half a century. Surprisingly, there are authors who believe that the approach of data protection by design is a much more recent development of the past ten years. This truncated historical understanding has resultantly led to some misinterpretation of Art. 25 GDPR. For instance, while Waldmann rightly recognised the need for differentiating between data protection and security, he asserts that Art. 25 GDPR is wholly insubstantial on its own, and only serves to remind data controllers to comply with all other GDPR provisions.[16] This reasoning points to a lack of conceptual understanding behind the methodological structure of Art. 25 GDPR.[17]

One reason for this dismissal might be the ambiguous object and concept of protection of the fundamental right to data protection in Article 8 ECFR, upon which the GDPR is "in particular" based (see the wording in Art. 1 sect. 2 GDPR). Some authors have already sketched out a historical reason for this methodological lack and proposed a solution in respect to other fundamental rights (esp. the right to private life in Article 7 ECFR). One of the main issues at stake in the discussion is the fact that as our society becomes increasingly digital, more and more social interaction relies on the processing of personal data. This in turn leads to an ever wider scope of application of the fundamental right to data protection, displacing other fundamental rights (which used to apply in the non-digital world). The question is therefore how to clarify the relationship between these fundamental rights so that we can also achieve in the digital world nuanced protection through a variety of applicable fundamental rights.[18] However, if we return to the level of ordinary law, another reason for further misinterpretations and misconceptions is the confused legalese of Art. 25 GDPR. The onslaught of words in Art. 25 GDPR makes comprehension extremely difficult (in contrast to the relatively clear wording of § 6 of the German Federal Data Protection Law from 1978). Art. 25 sect. 1 GDPR states:

> *"Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects."*

However, buried under this onslaught of words, there is a fundamental structure provided by Art. 25 GDPR.[19] Another contribution has examined in more detail how the methodical risk assessment in Art. 25 GDPR interacts with the legal principles under Article 5 GDPR, and how these principles are further specified by all the rest of the legal rules, in particular, under Artt. 6 et seq. GDPR.[20] In this contribution, we will thus provide only a brief summary of the normative scheme behind Art. 25 GDPR:

To begin, it is important to recognise that the whole GDPR follows, on an abstract-general level, a risk-based approach since it protects all fundamental rights (see Art. 1 sect. 2 GDPR) against the risks of personal data processing. The GDPR pro-

[14] See for example, the Privacy Icons Contest recently organised by the Italian Data Protection Authority, <https://www.gpdp.it/web/guest/temi/informativechiare> accessed 18th January 2022; see aso the more scientific approach, Habib, H., et al. (2021, May). Toggles, dollar signs, and triangles: How to (in) effectively convey privacy choices with icons and link texts. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-25).

[15] Regarding the first, see the sources at J. Pohle, 'Das Scheitern von Datenschutz by Design: Eine kurze Geschichte des Versagens' (2015) 2 FlfF-Kommunikation 32, p. 43; and regarding the latter, for instance 'Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation' 2016/679, 04 April 2017, 17/EN (Wp248rev.01) European Commission, pp. 8, 13, as well as in the Annex 2 - Criteria for an acceptable DPIA., p. 21: "origin, nature, particularity and severity of the risks are appreciated (cf. recital 84) or, more specifically, for each risk (illegitimate access, undesired modification, and disappearance of data) from the perspective of the data subject", criticized by M. von Grafenstein, 'How to Build Data-Driven Innovation Projects at Large With Data Protection by Design: A Scientific-Legal Data Protection Impact Assessment With Respect to a Hypothetical Smart City Scenario in Berlin' (2020) SSRN Scholarly Paper ID 3606140, p. 73, <https://www.hiig.de/publication/paper-data-protection-by-design-in-smart-cities/> accessed 23 March 2021.

[16] A. E. Waldman, 'Data Protection by Design? A Critique of Article 25 of the GDPR' (January 25 2021,) 53 Cornell International Law Journal.

[17] Cf. L. A. Bygrave, 'Data protection by Design and by Default: Deciphering the EU's Legislative Requirements' (2017) 1 OLR 105, p. 117.

[18] M. von Grafenstein, '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 Concepts of Risk Regulation', EDPL 04/2020.

[19] See, for example: on EDPB, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (adopted on 20 October 2020), version 2.0 .

[20] M. von Grafenstein, '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)', going to be published in EDPL 03/2021.

vides for such protection by establishing a set of legal principles, such as purpose limitation, lawfulness, data minimisation and transparency (Art. 5 GDPR), followed by an even bigger set of specifying legal rules.[21] In brief, the main difference between both types of legal instruments is that a legal principle sets a normative goal, giving the controller substantial room to manoeuvre: the controller should independently determine the best implementation of the legal principle according to the particularities of the specific case. This leeway makes legal principles useful instruments to regulate phenomena that are difficult to foresee such as risks (especially risks of data-driven innovation). Moreover, it allows for an evolving standard, such that law does not have to be redefined every other year - which arguably is not possible from a practical perspective. In contrast, legal rules are much more specific and consequently more restrictive; however, they also provide for greater legal certainty. For example, if a certain design element was to be defined by law for obtaining consent, this would increase certainty, but would force the legislator to revisit the issue once potentially better alternatives have arrived or run into the danger of holding back innovation and even work against effective data protection. The GDPR legislator therefore combines both legal instruments (so that they can complement each other) and requires, in addition, the controller to implement all these provisions into the processing design according to another, now individual-specific risk-assessment, pursuant to Art. 25 GDPR. That being so, the controller can mitigate the likelihood and severity of risks caused by its processing to the data subject's fundamental rights by implementing the legal principles and rules into the processing design (including its technological and organisational context).[22]

Of course, there are many entry points for further legal discussion. One issue at hand is whether the controller can directly refer to the fundamental right in concern—to adjust its measures according to the normative substance of this right at risk—or whether the controller may limit its attention to the "compliance risk" of not sufficiently implementing the legal principles and legal rules. In the second scenario, the correct implementation of legal principles and rules may protect the data subject's fundamental rights, but only in an indirect manner, without providing further guidance on how to design the data processing itself. Regardless, in order to assess the risks—whether they relate directly to fundamental rights or not—the controller may refer to a set of analytical terms: the nature, scope, context and purposes of the processing. All these terms are different analytical tools to carve out the particularities of the risks to one or more fundamental rights.

While Art. 25 GDPR is a complicated provision, one can carve out a rudimentary structure from the legalese.[23] It is apparent that the provision demands (a) the result of effective protection against the risks of data processing to the data subject's fundamental rights by (b) implementing a set of legal principles and rules into (c) the processing design (including its technological and organisational context). It is also apparent to some extent how one must assess the risks—this significantly aids in the allocation of protection efforts, supports effective protection, and reduces unnecessary regulatory burden. Additionally, the controller has a rich set of legal principles and legal rules at their disposal, legislative tools which may provide increased structure in the implementation phase for this kind of risk protection (not to speak of the regulatory burden that such repetitive protection systems cause).[24] This is the basic normative structure that Art. 25 sect. 1 GDPR provides for in its implementation of data protection by design. However, what remains unclear is how one may assess whether the implemented protection is effective. To answer this question, a clear methodology for assessing the effectiveness of said protection and, therefore, whether the controller is in compliance with Art. 25 is required. Thus far, even renowned experts have overlooked the need for establishing clear methodologies in assessing effective protection.[25]

1.3. *The dependency on interdisciplinary methods as a catalyst for better data protection*

In our opinion, Art. 25 GDPR's focus on the effects of protection measures is what makes it a groundbreaking piece of legislation. While Art. 25 narrows down each individual element of the complicated normative structure of the data protection by design approach (in contrast, for example, to the privacy by design approach where these elements remain rather vague),[26] the effectiveness requirement challenges classic legal reasoning because it requires lawyers to rely on methods beyond their own methodical scope. We believe that this methodological paradigm shift within a law, i.e. from disciplinary legal reasoning towards an explicit interdisciplinary approach, has not yet been understood to its fullest extent in the legal debate, despite its frequent occurrence, usage, and the fact that in some areas, an interdisciplinary approach has long been practised as a matter of course. One such example can be seen in the Guidelines 4/2019 on Article 25 Data Protection by Design and

[21] M. Eifert, Regulierungsstrategien (Regulation Strategies), in: Wolfgang Hoffmann-Riem, Eberhard Schmidt-Aßmann, Andreas Voßkuhle (eds.), Grundlagen des Verwaltungsrechts – Band I „Methoden – Maßstäbe – Aufgaben – Organisation" (C.H. Beck, 2nd ed., 2012), cip. 13 to 15; focusing on privacy-related principles, W. J. Maxwell, Principles-based regulation of personal data: the case of 'fair processing, in: International Data Privacy Law, 205–216, 5 (3) (2015), referring to Julia Black, Forms and Paradoxes of Principles Based Regulation', in: Capital Markets Law Journal, 3 (4), 425-457 (2008).

[22] M. von Grafenstein, '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)', going to be published in EDPL 03/2021.

[23] Cf. L. A. Bygrave, 'The Oxford Handbook Of Law, Regulation And Technology' (2017, 1st edn., Oxford University Press), p. 5.

[24] M. von Grafenstein, '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)', going to be published in EDPL 02/2021.

[25] I. Rubinstein and N. Good, 'Regulating Privacy by Design, Berkeley Technology' (2011) 26 Law Journal 3, pp. 1409–1456; 'The Trouble with Article 25 (and How to Fix It): The Future of Data Protection by Design and Default' (September 30 2019,), International Data Privacy Law (2020), Chapter ARTICLE 25 AND ITS DEFICIENCIES.

[26] A. Cavoukian, 'Privacy by design: The Definitive Workshop, A foreword by Ann Cavoukian, Ph.D' (2010) 3 Identity in the Information Society, pp. 247-251.

by Default, in which the EDPB states that the controller must define key indicators to measure and demonstrate the effectiveness of its implemented means: this includes, but is not limited to quantitative metrics and qualitative methods such as 'evaluations of performance, use of grading scales, or expert assessments'.[27] Similarly, in its transparency guidelines, the EDPB outlines that 'appropriate measures will need to be assessed in light of the product/service user experience' and that 'data controllers may wish to trial different modalities by way of user testing (e.g. hall tests) to seek feedback'.[28] Testing the effectiveness and intelligibility of the provided measures through qualitative and/or quantitative empirical measures is not something that data protection lawyers have been trained in thus far, and arguably do not have to. However, lawyers must be able to recognise the limitations of their own discipline, i.e. their own legal concepts and methods, and thus the interface with other disciplines that may be required for the full application of the law. If the application of the law requires such collaboration with other disciplines, this also requires that lawyers understand the concepts and methods of those other disciplines, at least in their basic approach. This basic understanding of one's own limitations and the capabilities and functioning of the other discipline applies not only to lawyers, of course, but also to the collaborators from the other discipline. The following considerations (esp. the second part of this article) will focus on this interface, i.e. how both disciplines may synchronize its concepts and methods in order to be able to answer the common research question.

Since this is a contribution to a legal journal, we will particularly address the legal concerns that often arise when empirical research results are taken into account in the legal doctrine. First of all, it is important to note that we do not say that the views of subjects asked in a pedestrian area could substitute for the normative standard as it has been expressed by the democratically elected legislator in the GDPR, even in the case of a broadly representative study.[29] Instead, it is our view that once the normative standard (and its basic assessment structure) is set, methods from outside the law may be better suited to assess whether the implemented protection measures are *effective* in matching said standard. With regard to evaluating the effectiveness of user interface design, the research field of User Experience Design has a lot to offer and a longstanding history of evaluating exactly such factors (e.g. effectiveness, but also transparency and controllability). This interdisciplinarity is no longer a rare exception in the legal sciences, but well established, for example in the field of IT security: The technical discourse on the effectiveness of encryption technologies is taken into account when assessing the effective application of the confidentiality principle quite naturally. The same goes for findings from the technical anonymisation discourse when trying to assess the effective implementation of the data minimisation principle. Still, there is a considerable number of lawyers commenting on the GDPR who nevertheless brush aside the use of empirical methods, namely user experience design methods, with a single sentence,[30] if it is in fact considered at all (most do not).[31] As a result, it is our argument that those who call themselves data protection law experts are in no position to discount the findings of other disciplines and their respective methodologies when they have the potential to support assessments of legal efficacy. This also applies to protection measures whose effectiveness depends on factors such as usability and utility, factors which are facilitated primarily by user experience designers rather than mathematicians or informatic engineers.[32]

Another issue at hand is how these findings are adopted into the legal discourse. The fact that this knowledge transfer can be accompanied by "translation barriers", or more precisely (as we have seen in the instance of anonymisation debates), that the same terminology used in different disciplines hides the fact that both disciplines actually refer to different concepts.[33] How these translation challenges are addressed and, more importantly, how potential connections across disciplines are formulated are crucial elements to successful interdisciplinary research. A decisive prerequisite to make this work is to recognise that the concepts, methods and/or findings can not, in most cases, be directly transferable from one discipline into another. Instead, each disciplinary discourse must carve out which elements it can take on from the other discipline and, more interestingly, how it can adapt its own concepts and methods to make this transfer even more effective in the process of reaching its respective research goals. In respect to legal practice, opening the law to concepts and methods borrowed from external research fields, as covered in Art. 25 GDPR, would not mean that adopted concepts and methodologies would be directly incorporated into the legal interpretation of the norm. Such direct adoption (e.g. of technical standards) does exist (e.g. in environmental law), but in such cases it is usually explicitly stated so in the law. If there is no such explicit reference to the direct application of non-legal standards, such as in the case of Art. 25, they are usually only taken into account as a secondary factor in

[27] EDPB, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (adopted on 20 October 2020), version 2.0, cip. 7.

[28] Art. 29 Data Protection Working Party, 'Guidelines on Transparency under Regulation' 2016/679, April 2017, 17/EN (Wp248rev.01), p. 13.

[29] Cf. A. Voßkuhle, 'Grundlagen des Verwaltungsrechts', § 1 Neue Verwaltungsrechtswissenschaft (2012, 2nd edn., C.H.Beck), pp. 1-65.

[30] J. Moser, 'Kommentar-Datenschutz-Grundverordnung' (2017, 1st edn., Bundesanzeiger Verlag), cip. 33.

[31] At least mentioning quantitative or qualitative indicators, EDPB 'Guidelines on on Article 25 Data Protection by Design and by Default' (adopted on 20 October 2020), version 2.0, cip. 16; the only source mentioning usability aspects and design methods at least with respect to IT security, at least in the German legal debate, is G. Hornung, M. Schallbruch, 'IT Security, Praxishandbuch' (2020, 1st edn.); see, in contrast, the literature regarding Art. 25 GDPR not mentioning UX design as a potential test method: S. Simits, G. Hornung, I. Spiecker, 'Datenschutzrecht: DSGVO mit BDSG, Großkommentar'(2019, 1st edn., Nomos); J. P. Albrecht, F. Jotzo, 'Das Neue Datenschutzrecht der EU' (2016, 1st edn., Nomos); B. P. Paal and D. A. Pauly, 'Datenschutz-Grundverordnung Bundesdatenschutzgesetz: DS-GVO BDSG' (2018, 2nd edn., C.H.Beck); H. A. Wolff et al., 'Datenschutzrecht in Bund Und Ländern' (2013, 1st edn., C.H.Beck).

[32] Cf. the panel '"Exploring the 'Design' in Privacy by design"' (2018) CPDP (24.01.2018 Brussels, Belgium).

[33] J. Hölzel, 'Differential Privacy and the GDPR' (2019) 5 European Data Protection Law Review, pp. 184-196.

the interpretation of the law.[34] Despite their indirect impact, the consequences of their inclusion may pose significant in terms of legal practice: interpreters of the law (e.g. a data protection authority or a legal court) cannot simply ignore methodologically-assured findings of the external discipline, since these practices and methodologies describe the factual situation on which the interpretation of the law is based.

In fact, this interdisciplinary inclusion in Article 25 fits into a larger development in the regulation discourse known as evidence-based policymaking.[35] Legislators are increasingly adopting an evidence-based approach, i.e. legislative measures based on empirical findings especially about the effectivity of the envisaged measures, for example, "to ensure the coherence, quality and delivery of policy, legislation and operations across policy areas and (…) departments", as the Secretariat General of the European Commission has explicitly set out.[36] In this debate, the increased rationalisation by evidence-based regulation has been discussed for quite some time now; similarly, the possible pitfalls of this approach, (including increased complexity in regards to the effects of regulation instruments in legal reasoning) have been debated.[37] In the rest of this article, we will return to these issues and offer potential solutions to the aforementioned pitfalls. However, in regards to the increase of complexity, it should be noted that Art. 25 GDPR at least makes clear that the data controller may take the implementation costs into account. Moreover, the compliance effort does not have to be repeated over and over again: once the effectiveness of a measure has been determined with respect to a certain type of risk and established in the market (especially in the case of so-called state of the art solutions),[38] it can, in principle, be adopted by others without any further methodological (e.g. empirical) effort. Only if the superior effectiveness of another measure or implementation path is to be proven, the non-legal (e.g. empirical) backing will be required again. In fact, it may be that a prevailing *legal* opinion on an (only seemingly effective) implementation of a GDPR requirement is even weakened because the results of another discipline suggest a different (actually more effective) implementation. Accordingly, an essential task of the legal discipline is to determine to what extent such results have to be taken into account (and thus also to what extent a state of the art solution in one context is transferable to the next). In any case, making the implementation of laws more effective means taking the results and methods of other disciplines into account. However, relying on non-legal methods to test the effectiveness of data protection by design requires the legislator to exercise caution in determining the measures themselves.[39] Consequently, only methodologically-robust research practices should be used to provide reliable guidance in determining which legal measures are effective, and to what extent these measures live up to the standard of being so-called 'state of the art'.[40]

2. Efficacy through UX-design: a case study on effective purpose specification

In this chapter, we will first carve out the current state of legal research regarding the normative aim and application of the principle of purpose limitation, especially purpose specification as its first component. Having clarified the corresponding legal research gap, we will then formulate our specific research question and explain in more detail why we consider the research disciplines of User Experience Design (UXD) as particularly suited to answer it. Turning to the point of view from UXD research, we will highlight some aspects where its research strands focusing on privacy can – now vice versa – take over certain findings from the data protection law debate.

Before jumping into the issue, we would like to provide the reader with a preparatory note: As often the case when combining disciplines, we would like to start off with a brief clarification of terminology, and the term "unfavourable use": In UXD research, the use of data that the data subjects would like to prevent is typically perceived as a 'risk' to privacy. Such risks include everything that users perceive to potentially interfere with their privacy, be it e.g. knowledge that others acquire about them and/or actions that are based on this knowledge and may be used against them. From a pragmatic point of view, perceived invasions of privacy – even if they do not objectively exist – are real in that they may steer behaviour because they are perceived as real by data subjects.

In the area of data protection law, however, the concept of risk is a complex matter and is heavily discussed, in particular, in terms of its relationship to other terms such as impact, threats and consequences. Therefore, throughout this paper, we do not refer to the term risk as used in the legal data protection debate, but stay closer to the user perspective. For clarification, we therefore mostly use the term 'unfavourable data use' (and if we should use the term 'risk', we put it into quotation marks). In this understanding, we follow the proposition

[34] M. Eifert, 'Grundlagen des Verwaltungsrechts', § 19 Regulierungsstrategien (2006,1st edn.) pp. 1237-1311.

[35] R. van Bavel, F. J. Dessart, 'The Case for Qualitative Methods in Behavioural Studies for EU Policy-Making' (27 April 2018) EU Science Hub, European Commission (2018) <https://ec.europa.eu/jrc/en/publication/eur-scientific.and-technical-research-reports/case-qualitative-methods-behavioural-studies-eu-policy-making> accessed 23 March 2021.

[36] Secretariat-general of the EU Commission, Strategic Plan 2020-2024, p. 4 <https://ec.europa.eu/info/sites/default/files/strategic_plan_sg_2020-2024_revised.pdf> accessed 18th January 2022.

[37] A. Voßkuhle, 'Grundlagen des Verwaltungsrechts', § 1 Neue Verwaltungsrechtswissenschaft (2012, 2nd edn., C.H.Beck), pp. 1-65.

[38] EDPB, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (adopted on 20 October 2020), version 2.0, cip. 18 et seqq. ; M.von Grafenstein, 'Co-Regulation and the Competitive Advantage in the GDPR: Data Protection Certification Mechanism, Codes Of Conduct and the "State Of The Art" of Data Protection-By-Design' (2019) SSRN Scholarly Paper ID 3336990 <https://papers.ssrn.com/abstract=3336990> accessed 23 March 2021.

[39] Cf. I. Rubinstein an N. Good, 'Regulating Privacy by Design, Berkeley Technology' (2011) 26 Law Journal 3, pp. 1409–1456; 'The Trouble with Article 25 (and How to Fix It): The Future of Data Protection by Design and Default' (September 30 2019,), International Data Privacy Law (2020), Chapter ARTICLE 25 AND ITS DEFICIENCIES.

[40] EDPB Work Program 2019/2020 (12 February 2019) <https://edpb.europa.eu/sites/edpb/files/files/file1/edpb- 2019-02-12plen-2.1edpb_work_program_en.pdf> accessed 12 February 2019.

of the Article 29 Working Group that the principle of purpose limitation should limit 'data use'.[41]

2.1. *Assessing the normative objective of purpose specification*

This contribution mostly focuses on the purpose limitation principle's primary component: purpose specification. The principle is established under Art. 5 sect. 1 lit. b) GDPR, which states that "personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes." On this basis, Section 1 lit. c) of Articles 13 and 14 GDPR require the controller to "provide the data subject with the purposes of the processing for which the personal data are intended" and, according to article 12 sect. 1 sent. 1 GDPR, take "appropriate measures to provide [… this information…] to the data subject in a concise, transparent, intelligible and easily accessible form." Thus, clarifying how processing purposes should be specified to be effective, at least from a data subject's point of view, is a necessary prerequisite to also implement the legal provisions on transparency appropriately. Last but not least, Article 25 sect. 1 GDPR requires the controller to implement the data protection principles, such as purpose specification and transparency, as well as the aforementioned legal rules through appropriate measures in an *effective* manner. In more colloquial terms, this means that the controller must specify the processing purpose and make this purpose explicit (i.e. transparent) to the data subject in an effective manner and intelligible form through appropriate measures. As noted, such an effectiveness assessment requires us first to clarify the purpose limitation principle's regulatory objective. The determination of this objective can then inform the process of selecting and applying an appropriate method which ensures the controller implements this principle effectively.

2.1.1. *What should purpose specification actually make transparent and controllable?*

Promoters of the purpose limitation principle argue that the principle protects data subjects by contributing to transparency, legal certainty, and predictability since purpose limitation "prevent[s] the use of individuals' personal data in a way (or for further purposes) that they might find unexpected, inappropriate or otherwise objectionable."[42] While this regulatory objective is undisputed, on both the practical and theoretical level it remains unclear how the controller can achieve this goal. In this context, many discussions refer to the second component of the purpose limitation principle, which obliges the controller to limit subsequent processing to the original purpose. According to the GDPR, data may not be processed in a way that is incompatible with the original purpose. Critics stress that limiting subsequent processing to original purpose would contradict the needs of our modern "information society": they argue that no one is capable of foreseeing all purposes for which data might be needed in the future.[43] However, only a handful of critics point out (or are even aware of) an ambiguity preceding this criticism. At first glance, the first component of the principle of purpose limitation appears straightforward, only requiring the controller to specify the purpose. However, it remains unclear *how* exactly a controller must specify the purpose or, conversely, *how* broadly the controller *may* specify the purpose. An answer to these questions is crucial because it determines the limits of the use of the data in question. On the one hand, a broadly defined purpose, e.g. to earn money, does not necessarily restrict the controller's room for manoeuvre.[44] In this case, criticism of the second component of the purpose limitation principle would be in vain. On the other hand, such a broadly-defined purpose is problematic since the purpose preconditions further legal requirements: for instance, in the assessment of whether the processed data is necessary, adequate or relevant, what legal basis the controller can invoke and which safeguards for the data subject must apply.[45] If the purpose is specified too broadly, it can hardly serve as a precise starting point to apply all these legal provisions, in other words: to implement legal requirements effectively. For instance, what must the controller actually make transparent by specifying its processing purpose? What should data subjects actually control, particularly by consenting—or not—to such a purpose? In short, what is the precise reference point for assessing legal compliance?

2.1.2. *Various approaches to specify processing purposes (or rejecting the approach in full)*

Many authors are therefore recommending that a purpose should not be too broadly defined.[46] For example, the European Data Protection Board (EDPB) argues that "marketing purposes", "IT-security purposes" or "future research" would be considered too broad if no further information on the processing circumstances were added.[47] However, the EDPB does not actually provide appropriate examples of a well-defined purpose specification. Also, the exact reason for a controller to specify the purpose to a certain degree of detail remains vague: The EDPB only states that the purposes must be specific

[41] Art. 29 Data Protection Working Party, 'Opinion 03/2013 on Purpose Limitation' (WP 203, 02 April 2013), p. 11.

[42] Art. 29 Data Protection Working Party, 'Opinion 03/2013 on Purpose Limitation' (WP 203, 02 April 2013), p. 11.

[43] L. Moerel, C. Prins, 'Privacy for the Homo Digitalis: Proposal for a New Regulatory Framework for Data Protection in the Light of Big Data and the Internet of Things' (2016) SSRN Scholarly Paper ID 2784123, <https://ssrn.com/abstract=2784123 or http://dx.doi.org/10.2139/ssrn.2784123> accessed 23 March 2021; J. Soria-Comas, J. Domingo-Ferrer, 'Big Data Privacy: Challenges to Privacy Principles and Models' (2016) 1 Data Science and Engineering 21.

[44] N. Forgó, T. Krügel, S. Rapp, 'Zwecksetzung und informationelle Gewaltenteilung: Ein Beitrag zu einem datenschutzgerechten E-Government' (2006, 1st edn., Nomos); M. Kring, 'Big Data und der Grundsatz der Zweckbindung, Informatik' (2014), pp. 551-562; V. Mehde, 'Handbuch der europäischen Grundrechte' (2020, 2nd edn., C.H.Beck).

[45] EDPB, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (adopted on 20 October 2020), version 2.0.

[46] B. P. Paal et al., 'Datenschutz-Grundverordnung Bundesdatenschutzgesetz' (2018, 2nd edn., C.H.Beck), cip. 27; S. Simitis, 'Bundesdatenschutzgesetz' (2014, 8th edn., Nomos), cip. 78; E. Ehmann et al., 'DS-GVO: Datenschutz-Grundverordnung' (2018, 2nd edn., C.H.Beck), cip. 14.

[47] Art. 29 Data Protection Working Party, 'Opinion 03/2013 on Purpose Limitation' (WP 203, 02 April 2013), p. 16.

enough to enable data subjects to foresee how the data will be used and to assess whether or not this is appropriate. However, it remains unclear how the controller should actually be able to assess what kind of data use a data subject considers relevant and appropriate. Unfortunately, very few authors tackle this question in greater depth. Mehde, at least, argues for a consideration of the consequences on data subjects.[48] However, it is worth questioning how data controllers should know what consequences are relevant from the data subjects' perspective.

The difficulty involved in the assessment of this perspective may be the reason why other authors believe that the controller may maintain their own perspective rather than incorporating the data subjects' point of view. Basin, Debois and Hildebrandt argue, for instance, that controllers can refer to factual business processes to specify their processing purposes.[49] At first glance, this offers a certain advantageous appeal, since business processes appear to be somewhat more objective than the data subjects' perspective. Both controllers and data subjects could therefore assess, in relation to these given processes, what data use can and must be expected. However, this assumption has two shortcomings: first, business processes are not objectively provided: they are the result of the controller's active decision to design its organization.[50] In support of this view, Albers considers the purpose specification as a process that bundles various data processing activities into a meaningful unit.[51] Similarly, the Technology Working Group (Arbeitskreis Technik) of the German Data Protection Agencies (Konferenz der unabhängigen Datenschutzbehörden des Bundes und der Länder) argues that business processes are designed to fulfil a specific purpose.[52] Thus, purpose specifications which fall under business process purposes are not objective; rather, they fulfil the controllers' subjective business goals. However, what the business purpose is and how this should be specified remains unanswered. More importantly – this is the second shortcoming – the inherent organisational sense to a business process does not necessarily constitute relevant or comprehensible information from the data subjects' perspective. Thus, even if the business processes were rendered transparent for them, this does not necessarily lead to meaningful information from their point of view. Therefore, it remains open *how* controllers should make their business processes transparent, or how to specify the purpose of these processes in order for data subjects to assess the level of relevance and appropriateness to their own situation.

This ambiguity on *how* the controller must specify its purpose is also where information scientist Nissenbaum begins her argument. Nissenbaum highlights that in the absence of any objective scheme for a reliable purpose specification, the purpose-based approach "constitute[s] a mere shell, formally defining relationships among the [privacy aka data protection] principles and laying out procedural steps to guide information flows", which in the end opens a "glaring loophole".[53] Consequently, Nissenbaum promotes her contextual integrity-approach as something better than the purpose-based approach.[54] According to her approach, informational contexts consist of informational norms which determine the "contextual integrity" of information. There are two aspects that make Nissenbaum's approach interesting for our contribution: On one hand, she promotes an empirical approach by clarifying that "*contexts are not formally defined constructs, but (…) are intended as abstract representations of social structures experienced in daily life. (…). In other words, the activity of fleshing out the details of particular types of contexts, such as education or health care, is more an exercise of discovery than of definition.*"[55] Thus, Nissenbaum argues in favour of providing a concept for applying empirical research methods to assess how data subjects want information about them to be used. On the other hand, her context-based approach is not as different from a purpose-based approach as she asserts. In fact, it is easy to understand the purpose-based approach as just another regulatory entry point to govern the use of personal data (aka information). Consequently, the processing purpose simply connects the context where personal data has been collected to a (potentially different) context in which the data is intended to be used. From that perspective, the purpose-based approach only shifts the focus of (regulatory) attention to the point *before* the data is used in a context, and before the contextual integrity of that information is violated. This shift is the consequence from the preventative risk regulatory approach in data protection law.[56] Thus, it is as reasonable to empirically assess the expected use of personal data with respect to the processing purpose as it is with respect to Nissenbaum's context-based approach. In principle, the concepts, methods and findings of both approaches can complement and build on each other.

[48] V. Mehde, 'Handbuch der europäischen Grundrechte' (2006, 1st edn., C.H.Beck), cip. 24.

[49] D. Basin, S. Debois, T. Hildebrandt, 'On Purpose and by Necessity: Compliance Under the GDPR', Financial Cryptography and Data Security (2018, Springer).

[50] R. Beckhard, 'Organization Development: Strategies and Models' (1969); V. Wulf, M. Rohde, 'Towards an Integrated Organization and Technology Development' (1995) Proceedings of the 1st conference on Designing interactive systems: processes, practices, methods, & techniques, Association for Computing Machinery.

[51] M. Albers, '§ 22 – Umgang Mit Personenbezogenen Informationen und Daten, Grundlagen des Verwaltungsrechts Informationsordnung – Verwaltungsverfahren – Handlungsformen', Grundlagen des Verwaltungsrechts (2012, 2nd edn., C.H.Beck), cip. 123.

[52] 'Konferenz der unabhängigen Datenschutzbehörden des Bundes und der Länder, The Standard Data Protection Model – A method for Data Protection advising and controlling on the basis of uniform protection goals' (2019) version 2.0b, p. 9.

[53] H. Nissenbaum, 'Respect for Context as a Benchmark for Privacy Online: What It Is and Isn't', Social Dimensions of Privacy: Interdisciplinary Perspectives (2015, Cambridge University Press), pp. 278-302.

[54] H. Nissenbaum, 'Respect for Context as a Benchmark', ibid., p. 292.

[55] H. Nissenbaum, 'Privacy in Context, Technology, Policy, and the Integrity of Social Life' (2020, Stanford University Press), p. 134.

[56] M. von Grafenstein, '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)', going to be published in EDPL 02/2021; M. von Grafenstein,' The Principle of Purpose Limitation in Data Protection Laws: The Risk-Based Approach, Principles, and Private Standards as Elements for Regulating … Innovationsforschung' (2018, 1st edn., Nomos), pp. 104, 105.

2.1.3. *Conclusions for our research question and inductive-empirical approach to answer it*

On the basis of the preceding summary of the debate, we formulate the following research question to demonstrate both our understanding of the purpose limitation principle, and more generally, of the data protection by design approach:

> *How should purposes be effectively specified from the data subject's perspective so that the purpose specification 'prevent[s] the use of individuals' personal data in a way (or for further purposes) that they might find unexpected, inappropriate or otherwise objectionable'?*[57]

We see two fundamentally different research approaches to answer this question. As a first research approach, one may proceed deductively by establishing an objective scheme for the data subjects' reasonable expectations of the data subjects. For instance, as Max Grafenstein argued in a former contribution, the variety of all data subjects' fundamental rights should be understood as such an objective scheme.[58] Accordingly, the purpose specification requirement serves to discover risks that the data processing might cause against one or more of the fundamental rights of the data subject. If the intended processing causes a specific risk to such a right, the controller must specify this risk and explicitly inform the data subject (and apply all necessary protection measures) accordingly. If the controller fails to do so, then he may not process the data in a way that presents a de facto risk to the data subject's rights. Thus, the canon of all fundamental rights (often further specified by ordinary law) provides an objective scheme for categorising the legally relevant consequences for the data subject and, as such, must be accepted and respected both from the data subjects' and controller's perspective. However, such an objective scheme still carries the risk of not being effective in informing laypersons about the data processing consequences. While it is perfectly possible to empirically test the effectiveness of such an objective scheme determined by the data subject's fundamental rights, it seems more promising to directly ask the data subjects instead, since their answers would provide independent empirical data. When it comes to the question of how to formulate purposes in a meaningful and effectively transparent manner for the data subjects themselves, it is striking that legal research (to the best of our knowledge) has not previously taken into account empirical findings on the data subjects' perspective.

Consequently, we have applied a more inductive and social constructivist approach. In the search for effective means of creating transparency and control of the processing, we argue in favour of shifting the perspective towards the data subjects themselves. More specifically, we ask them what they need in the formulation of processing purposes to feel effectively informed about how their data will be used.

2.2. *Choosing the empirical methodology of UXD research*

For evaluating the effective implementation of the law, the legal sciences do not have to change themselves nor invent new methods. Instead, there are several research areas to choose from that already have an established arsenal of methods. For example, methods from the behavioural sciences, psychology, or cognitive research are suitable for quantitative studies. But especially when it comes to the perception of and interaction with technical artifacts, the field of user experience design arguably is the most appropriate, versatile field. Research on UXD is interdisciplinary in two respects: First, it is not ideological but pragmatic on a methodological level: It uses quantitative test methods, for example, to test existing hypotheses, but is also able to explain complex phenomena, for example, through ethnomethodological methods or the study of the appropriation of technology and, thus, to develop research hypotheses, qualitatively. UXD has its origins in the field of office workstation ergonomics, where the term "usability" was initially used to measure the effectiveness, efficiency and satisfaction in the use of software.

Effectiveness of digital technology is therefore in a sense the core of the origin of UXD. UXD itself then emerged from the realization that (not only, but especially) in the private sphere it is about more than the concrete interaction with technology, but also about expectations and feelings beforehand and afterwards when examining acceptance. Second: It is interdisciplinary by nature, in seeking to combine viewpoints and expertise of different stakeholders in designing solutions. With the principle of purpose limitation having the task of making data processing activities transparent to and controllable by users, it is a perfect fit for a demonstration of the merits of UXD research for identifying and assessing effectiveness of implementations.

2.2.1. *The aim of UXD: increasing vs. re-balancing power asymmetries?*

To answer our research question, we combine legal reasoning with methods and concepts of UXD research. Beyond classical usability testing, which assesses the values of effectiveness, efficiency and satisfaction, UXD methods also aim at designing systems and processes that meet users' universal needs: this includes factors such as competence, stimulation, meaning, security, or autonomy.[59] Users should also be able to understand and use these systems and processes, find them useful, and in the best case, enjoy them.[60] UXD can be understood as a broader perspective on the traditional field of Usability Engineering, which previously focused on ISO 9241–11

[57] Art. 29 Data Protection Working Party, 'Opinion 03/2013 on Purpose Limitation' (WP 203, 02 April 2013), p.11.

[58] M. von Grafenstein, '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)', going to be published in EDPL 02/2021.

[59] M. Hassenzahl, S. Diefenbach, A. Göritz, 'Needs, Affect, and Interactive Products – Facets of User Experience' (2010) 22 Interacting with Computers 5, pp. 353-362; K. M. Sheldon, A. J. Elliot,'What Is Satisfying about Satisfying Events? Testing 10 Candidate Psychological Needs' (2001) 80 Journal of Personality and Social Psychology, pp. 325-339.

[60] F. Schaub, L. F. Cranor, Chapter 5: 'Technology, Academics, Policy – Usable and Useful Privacy Interfaces' (2020) IAPP, p. 189 et seqq.

and the concepts of effectiveness, efficiency and satisfaction. As such, UXD also involves the emotional states of users, extending beyond the goal of task completion. Additionally, it is more suited to address socio-technical issues such as managing privacy online. Distinguishing UXD research from Human Computer Interaction (HCI) nowadays arguably merely is a matter of the audience.[61] Since the aim of our research is to have an impact on how effectiveness of data protection by design measures are proven (esp. on how processing purposes are specified) in future everyday practice, we here prefer to use the term that might be more known in the legal debate: UXD.

The development of concepts and methods of UXD research have been a great success. Putting these methods to practice, UXD contributes to designing digital technologies in an intuitively usable and even pleasurable way: indeed, they have become indispensable companions in our daily life. However, the success of UXD as part of the digital economy also means, to put it bluntly, getting the user to click the "buy" button as seamlessly and quickly as possible. Likewise, the consumer is encouraged to provide as much personal information about themselves as possible. Given the increase of informational power asymmetries facilitated by the increased usability of technologies, it stands to reason that the same UXD methods can be successfully applied to more ethical value-oriented objectives such as those substantiated in data protection law. Of course, this observation is not meant to disparage the successes of UXD research in general. In fact, the community is keen on investigating the emergence of so-called "dark patterns" in interaction design.[62] Regarding the legal debate, cookie banner design has gained a lot of attention in recent discussion.[63] Likewise, specialised methodologies that address such value-oriented goals are not new to UXD, but have a long history of development as seen in the evolution of e.g. regulation by design,[64] and privacy by (UX) design.[65] In these debates, Urquhart's and Rodden's observation offers particular insight on what their field (i.e. UXD research) can add to the legal discussion on regulation and design. In their opinion, the classic discussion of regulation and design is still anchored in system theory, while UXD research has long shifted from a top-down approach "to more situated, nuanced understandings of the contexts of design".[66] This matches with our own observation within the data protection law debate, where there is a recognizable tendency to reduce the data protection by design approach to technical data and IT security measures. Assuming that data protection by design and IT security are regarded as separate issues, it is not as necessary for the data subject to be the centre of the design process in issues of IT security than of data protection. Indeed, some IT security measures must be usable and useful to be effective (e.g. password management). However, data protection measures, such as those established from the principles of purpose limitation and transparency, as well as the variety of intervention rights (i.e. data access, correction, and complementation, deletion, consent, withdrawal, objection, and so on) depend almost entirely on their usability and utility. This makes it necessary to place the data subjects more rigorously at the centre of the data protection by design process than it is necessary in the area of security by design.

In this vein, the UXD research community has already conducted considerable research on "usable privacy".[67] For example, there is plenty of research on poor design privacy policy,[68] the limits of "notice and consent"-structures,[69] the current implementations of cookie banners,[70] and tracking.[71] Additionally, for a variety of systems ranging from smartphones to embedded and networked devices, UXD research has looked into user demands in terms of system intelligibility,[72] aware-

[61] M. Hassenzahl, A. Monk, 'The Inference of Perceived Usability From Beauty' (2010) 25 Human–Computer Interaction 3, pp. 235-260.

[62] C. M. Gray et al., 'The Dark (Patterns) Side of UX Design' (2018) Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery.

[63] I. Sanchez-Rola et al., 'Can I Opt Out Yet? GDPR and the Global Illusion of Cookie Control' (2019) Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security, Association for Computing Machinery; R. Bornschein, L. Schmidt, E. Maier, 'The Effect of Consumers' Perceived Power and Risk in Digital Information Privacy: The Example of Cookie Notices' (2020) 39 Journal of Public Policy & Marketing 2, pp. 135-154.

[64] L. Urquhart, T. Rodden, 'A Legal Turn in Human Computer Interaction: Towards Regulation by Design for the Internet of Things' (2016) SSRN, <https://www.research.ed.ac.uk/en/publications/a-legal-turn-in-human-computer-interaction-towards-regulation-by- accessed 23 March 2021.

[65] F. Schaub, L. F. Cranor, Chapter 5: 'Technology, Academics, Policy – Usable and Useful Privacy Interfaces' (2020) IAPP, p. 189 et seqq.

[66] L. Urquhart, T. Rodden, ibid., p. 33, referring to Lessig, Black, Scott, and Murray.

[67] C. Brodie et al., 'Usable Security and Privacy: A Case Study of Developing Privacy Management Tools' (2005) 49 ACM 1.

[68] S. Norman, et al., 'The Usable Privacy Policy Project' (2013) Technical report CMU-ISR-13-119; C. A. Brodie, C-M Karat, J. Karat, 'An Empirical Study of Natural Language Parsing of Privacy Policy Rules Using the Sparcle Policy Workbench (2006) Proceedings of the Second Symposium on Usable Privacy and Security.

[69] A. Acquisti, I. Adjerid, L. Brandimarte, 'Gone in 15 Seconds: The Limits of Privacy Transparency and Control' (2013) 11 Security & Privacy IEEE, pp. 72-74; I. Bilogrevic, M. Ortlieb, '"If You Put All The Pieces Together…"' (2016) Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 5215-5227.

[70] I. Sanchez-Rola, et al., 'Can I Opt Out Yet? GDPR and the Global Illusion of Cookie Control' (2019) Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security, Association for Computing Machinery; R. Bornschein, L. Schmidt, E. Maier, 'The Effect of Consumers' Perceived Power and Risk in Digital Information Privacy: The Example of Cookie Notices' (2020) 39 Journal of Public Policy & Marketing 2, pp. 135-154.

[71] M. Degeling et al., 'We Value Your Privacy … Now Take Some Cookies: Measuring the GDPR's Impact on Web Privacy' (2019) Proceedings 2019 Network and Distributed System Security Symposium, NDSS; T. Jakobi et al., 'Web Tracking Under the New Data Protection Law: Design Potentials at the Intersection of Jurisprudence and HCI (2020) 19 i-com 1, pp. 31-45.

[72] V. Bellotti, K. Edwards, 'Intelligibility and Accountability: Human Considerations in Context-Aware Systems' (2001) 16 Human–Computer Interaction, pp. 193-212; T. Jakobi et al., 'Evolving Needs in IoT Control and Accountability: A Longitudinal Study on Smart Home Intelligibility' (2018) 2 Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 4, pp. 1-28.

ness about privacy implications,[73] as well as data literacy[74] as a means to cope with increasingly complex interrelations between data collection and its usage. Also, designing for transparency is a major theme in usable privacy research and it frequently mentions the risk-based approach as part of user practices to weigh the implications of data disclosure.[75]

2.2.2. *Methods for placing the user (i.e. data subject) at the centre of the design process*

In the field of UXD, the use of the user-centred design (UCD) process has long become a quasi-standard for the design of applications related to personal context. Examples range from home care, and family interactions around new media to energy management and sustainability.[76] Applying the methods and tools of a UCD process often aims at fostering a process of designing software that is, to varying extents, grounded in practice.[77] In UCD, users are the primary resource for understanding needs in system design (for instance: 'What is needed?' and 'What is possible?'). These questions are central to designing processing purposes in a meaningful way from a data subject's perspective. A user-centred approach to data protection is well-suited for designing viable purpose specifications which can incorporate legal compliance, organizational leeway for innovation, and transparency about the consequences for data subjects. Typically, such an approach attempts to assess the user experience of existing data protection solutions, analyse the conflict(s) between stakeholder interests, and identify viable compromises.[78]

Although there is a growing body of literature on mental models and design guidelines on a wide range of privacy and data protection issues in UXD, to the best of our knowledge, there remains a noticeable lack of empirical studies on user demands with respect to the formulation of processing purposes. Therefore, our research focuses on methods for generating new insights. To systematically identify design candidates that might be useful, qualitative methods are well-suited for this exploratory process. Qualitative methods make it possible to investigate phenomena and understand them in depth. To gain an accurate impression of how users interact with the research topic at hand, UXD researchers often adopt ethnomethodological methods, including observations 'in the wild', interviews, and workshops. Such qualitative insights may provide intimately personal descriptions of phenomena: the advantage these descriptions present allow us, for instance, to derive new design guidelines or apply theory-generation through techniques such as Grounded Theory.[79] It is important to notice that while qualitative studies still typically exhibit empirical saturation, (i.e. by which new phenomena no longer occur when more participants are recruited), they do not aim to generate representative findings.[80] Rather, qualitative work targets the identification of prevailing concepts and makes these newly generated hypotheses accessible to an iterative design process, in which users may evaluate prototypes for researchers to subsequently refine or evaluate them. Such an evaluation can take place at different levels of maturity, depending on the goal: for instance, an evaluation can focus on abstract concepts or concrete implementations of a system. In this stage, achieving a dataset that represents a wide variety of test samples is often the main goal, and evaluative studies are resultingly quantitative in nature. The success of UXD can be attributed to the triangulation of its various methodologies, resulting in new and reliable insights on how user demands can be met through the design of technology.

2.2.3. *Adjusting the UXD research design to our legal research question*

While we have argued that the legal debate can benefit from the use of UXD methods, we also put forth the argument that the converse is true: legal research aspects can and should be adapted to other research domains, such as UXD research, if it stands to benefit from its inclusion. We see three main reasons for why UXD research may benefit from taking legal-conceptual considerations into account. First: the legal domain can highlight open research questions to UXD research. As outlined in 2.1.3, the research question offers several interesting issues from a UXD perspective. In the last years, several studies started looking into the perceived unfavourable use of data as a design resource (especially in the realm of embedded and networked devices known as the Internet of Things).[81] These efforts focus on investigating potential design resources that increase user awareness of privacy implications when using services. They do not, however, take into account the principle of purpose limitation, or its first component of purpose specification as expressed in privacy poli-

[73] N. Gerber, B. Reinheimer, M. Volkamer, 'Investigating People's Privacy Risk Perception' (2019) Proceedings on Privacy Enhancing Technologies, pp. 267-288; T. Jakobi et al., 'It Is About What They Could Do with the Data: A User Perspective on Privacy in Smart Metering' (2019) 26 ACM Transactions on Computer-Human Interaction 1, pp. 1-44.

[74] P. Tolmie et al., '"This Has to Be the Cats"' (2016) Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, pp. 491-502.

[75] T. Jakobi et al., 'It Is About What They Could Do with the Data: A User Perspective on Privacy in Smart Metering' (2019) 26 ACM Transactions on Computer-Human Interaction 1, pp. 1-44; F. Díaz, E. Nicolás et al., 'Preventative Nudges: Introducing Risk Cues for Supporting Online Self-Disclosure Decisions' (2020), p. 399; V. Garg et al., 'Risk Communication Design: Video vs. Text' (2012) 7384 Privacy Enhancing Technologies, pp. 279-298.

[76] A. Crabtree, T. Rodden, 'Domestic Routines and Design for the Home' (2004) 13 CSCW, pp. 191-220; L. Palen, S. Aaløkke, 'Of Pill Boxes and Piano Benches: "Home-Made" Methods for Managing Medication' (2006) 10 Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work.

[77] V. Wulf et al., 'Practice-Based Computing: Empirically Grounded Conceptualizations Derived from Design Case Studies' (2015) Designing Socially Embedded Technologies in the Real-World, <https://doi.org/10.1007/978-1-4471-6720-4_7> accessed 23 March 2021.

[78] C. Brodie, et al. 'Usable Security and Privacy: A Case Study of Developing Privacy Management Tools' (2005) 49 ACM 1, pp. 56-57.

[79] B. Glaser, A. Strauss, E. Strutzel, 'The Discovery of Grounded Theory: Strategies for Qualitative Research' (1968) 17 Nursing Research 4, p. 364.

[80] I. T. Coyne, 'Sampling in Qualitative Research. Purposeful and Theoretical Sampling; Merging or Clear Boundaries?' (1997) 26 Journal of Advanced Nursing 3, pp. 623-630.

[81] . Jakobi et al.,'It Is About What They Could Do with the Data: A User Perspective on Privacy in Smart Metering' (2019) 26 ACM Transactions on Computer-Human Interaction 1, pp. 1-44.

cies.[82] Similarly, while the work on usable privacy policies in UXD research has found that privacy policies lack readability and comprehensibility, said research does not seek to reformulate purposes to make them more meaningful; rather, they accept the wording used as given.[83] Part of our contribution for the UXD research community therefore lies in highlighting that purpose specifications are a design resource in which data subjects are likely to have justified interests. We therefore examine the interests of data subjects in maintaining transparency of the consequences of data disclosure, focussing explicitly on the question of how processing purposes *should* be specified to limit 'unfavourable uses of their data'. To our point, the mere identification of our research question was an important first contribution from the data protection law debate to UXD privacy research, first and foremost because we highlight the importance of the regulatory objective of the purpose limitation principle.

A second reason for why UXD research may benefit from taking legal-conceptual considerations into account is to keep up with the discussion surrounding normative standards. As mentioned, breaking down the aspects of a normative standard in sufficient depth and detail is a particular strength of legal research: thus, it can provide a rich set of details of normative expectations on certain design artefacts in general. This normative analytical-conceptual work has a long tradition in law. At the same time, UXD privacy research limits itself to the individual user expectation level (of e.g. managing privacy), neglecting insights from the normative experts debate. While UXD research focuses on the expectations of *users*, one should not discount decades of research conducted by data protection and privacy experts who have sought to closely define the same or, at least, similar expectations, and upon which current laws are built. As a result, we sought to enrich the user-centred design process with elements typical to the legal debates surrounding issues of data protection and data privacy.

We decided to bridge potential "knowledge gaps"' between the users' and experts' perspectives by confronting them with their respective points of view. In the case of users, we presented them analytical ideas from data protection and privacy debates during an early exploratory design process. We usually started asking the users for their opinions without giving them any further information and thus received their "unfiltered" views. Later on, we introduced arguments from the expert debate into the discussion, and asked users for their take on these arguments.

However, legal-conceptual considerations also played a role in shaping the design process itself. The main reason for this resulted from our understanding of the implicit normalizing power of categorization, which we wanted to stay out of as much as possible in order to avoid authors' bias. This is best understood by briefly describing the consequences of this thought on the structure of our study design itself: To gain a thorough understanding of how users (i.e. data subjects) expect how their data might be used, and to use this understanding as a basis for an effective specification of processing purposes, we first conducted focus group interviews which concentrated not only on the generation of user-perceived unfavourable data uses but also on the respective categorization of these unfavourable data uses. Categorizing the variety of user-perceived 'unfavourable data uses' is a necessary step to deliver a consistent reference scheme for more effective specification of purposes. Given the need to set a consistent standard in law, the classification scheme serves as a powerful tool to normalize such standards. Since the categories formed by the user groups were extremely wide-ranging and disparate, we therefore considered including a third perspective. In order to prevent authors' bias, we did not bring in a "third perspective" by ourselves but consulted with external experts and asked them to categorize the unfavourable data uses collected in the user workshops. Including this third perspective allowed us to limit our own influence on the categorization process as well as bridge the aforementioned "knowledge gap" in now confronting experts with the data subjects' perspective. On this basis, we were able to compare the categories of both parties and looked for potential differences and similarities. Only in this stage, thus, did we bring our own perspective into the process, analysing the various categorization schemes and merging them into a final set in order to demonstrate common ground between expert interviewees and workshop participants.

The previous point leads us to a final reason for why UXD research may benefit from taking legal-conceptual considerations into account. In order to be able to be transferred into the logic of legal argumentation, the findings from our user-centred design research process must fit to the inherent standards set forth by law. Aiming at providing legal certainty, each law implies a scheme that is conducive to consistent reproducibility in future cases; additionally, the law must indicate whether future cases fall under the provided scheme. For example, if data subjects say that the purpose of 'IT security' must clearly indicate that the collected data may be used against them in a legal cyber security proceeding (e.g. if their home router has been hijacked by a botnet for denial of service attacks against another website), the idea behind this statement must also demonstrate why this purpose does

[82] N. Gerber, B. Reinheimer, M. Volkamer, 'Investigating People's Privacy Risk Perception' (2019) Proceedings on Privacy Enhancing Technologies, pp. 267-288; T. Jakobi et al., 'Evolving Needs in IoT Control and Accountability: A Longitudinal Study on Smart Home Intelligibility' (2018) 2 Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 4, pp. 1-28; S. Karwatzki, M.Trenz, D.Veit, 'Yes, Firms Have My Data but What Does IT Matter? Measuring Privacy Risks' (2018) ECIS; A. Rao, F. Schaub, N. Sadeh, 'What Do They Know about Me? Contents and Concerns of Online Behavioral Profiles' (2015) PASSAT 14: Sixth ASE International Conference on Privacy, Security, Risk and Trust. P. Tolmie et al., '"This Has to Be the Cats"' (2016) Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, pp. 491-502.

[83] A. M. McDonald, L. F. Cranor, 'The Cost of Reading Privacy Policies' (2008) 4 A Journal of Law and Policy for the Information Society 3 , pp. 543-568; J. R. Reidenberg et al., 'Disagreeable Privacy Policies: Mismatches between Meaning and Users' Understanding' (2014) SSRN Scholarly Paper ID 2418297 <http://papers.ssrn.com/abstract=2418297> accessed 23 March 2021; R. I. Singh, M. Sumeeth, J. Miller, 'Evaluating the Readability of Privacy Policies in Mobile Environments' (2011) 3 International Journal of Mobile Human Computer Interaction 1, pp. 55-78 ; N. Steinfeld, 'I Agree to the Terms and Conditions: (How) Do Users Read Privacy Policies Online? An Eye-Tracking Experiment' (2016) 55 Computers in Human Behavior, pp. 992-1000.

not have to make other consequences apparent (e.g. price-discrimination). Thus, if a scheme proposed by users is not consistent, it is difficult to use it for legal reasoning. This is another reason why the perspective of the users should by no means be regarded as the solution in itself; rather, their opinion should be weighed with respect to other stakeholder parties, and in particular with respect to the normative standards set forth by law. Beyond the standards of legal requirements, however, it has been demonstrated that understanding the phenomenon of control and transparency from a user perspective is a promising way of designing information systems in a more usable and useful way for a wide variety of contexts.[84]

3. Qualitative research findings

In the following, we present in more detail our qualitative research design and the research findings that are most relevant for our research question. As said, our research question is how purposes should be effectively specified from the data subject's perspective so that the purpose specification 'prevent[s] the use of individuals' personal data in a way (or for further purposes) that they might find unexpected, inappropriate or otherwise objectionable'? To answer this question, we wanted to find out which 'data use' categories users consider relevant to be more clearly included or excluded by processing purposes. Only on this basis is it possible to determine whether current purpose specifications meet the regulatory aim of the principle of purpose limitation – and if not, how they should be specified instead.

3.1. *Conducting the user workshops*

To this aim, we first defined three use cases in which data subjects have to manage the disclosure of personal data. To include a wide range of cases, we looked for both established and new technologies, featuring different modes of interaction: first, the well-established use case of web-browsing. In the second use case, we turned our attention to voice assistants as an example for the growing sector of new, market-driven Internet of Things (IoT) devices and services. Our third use case referred to connected cars and, thus, to pre-existing devices modernized to now connect to the internet. These use cases were the starting point for the discussions in our user workshops. Overall, we had 42 participants divided into nine focus groups, each of which was assigned one of the three use cases (so each use case was discussed in three groups). We designed our workshops in a way that let participants articulate, discuss and categorize perceived privacy risks in terms of unfavourable use of data outlined above. As previously emphasised, the use of the term 'risk' in our user workshops was an obvious choice, as both legal and UXD research on data protection and privacy, the risk-based approach is very present.

[84] C. Abras, D. Maloney-Krichmar, J. Preece, 'User-Centered Design, W. Encyclopedia of Human-Computer Interaction' (2004), p. 445.

3.1.1. *User acquisition and sample*

For the user workshops, we recruited our user sample through a self-hosted website, where interested people were asked to sign up, leaving their email addresses for further information. To attract potential participants to visit the website, we placed information about the study in the local press, on radio stations and in facebook ads. Participants were not offered any compensation for participation. Instead, the only incentive we offered was a free tour of a new interdisciplinary research laboratory in the university buildings, which was not yet open to the public, as well as more information about the project itself. Through e-mail exchanges, we asked participants to provide basic information on their demographics and technological experience in order to compile a well-mixed sample. In addition, we held three workshops with students from one of the partner universities. Per se, we tried to include all persons interested, but were unable to do so, simply for reasons of finding an appointment acceptable for all. Optimizing scheduling to minimize loss, our overall sample consisted of 42 participants. All participants were of German nationality and of legal age (ranging from 20 to 60 years). Self-reported tech-savviness varied: While all participants told us to frequently use the web, some participants reported e.g. not having used smart speakers and not knowing how to use connected mobility services nor how they work. With regard to legal expertise, there were two persons with a legal background among the participants and a data protection officer.

3.1.2. *Design of the workshops*

Our workshop design was very similar for each use case outlined above. First, we welcomed participants, introduced each other and outlined what to expect from the next two hours. We also informed them about the background of the project and that our study aimed at understanding the design of purpose limitation and, on this basis, potential support of privacy icons in this matter from different stakeholder perspectives.

1. Our first interest was to gain a common understanding of the technology in question. Therefore we gave room for individual views and had also prepared promotional videos of products for voice assistants and connected mobility services. Since the discussion on each topic proved to be sufficient to bring all participants up to a similar level, we did not show the promotional videos in the end.
2. In order to give participants the opportunity to reflect on their relationship with the service in question, we asked participants to name use cases that they know of, that they themselves use, or that they might adopt in the future. These use cases were collected on a poster that everyone could refer to.
3. We turned to unfavourable use and asked for 'anti use-cases', i.e. situations or contexts in which users would explicitly not use or have not used the service to fulfil a need. For example, we asked for a context in which the participants would not like to use the web, although it would be possible to use it to reach their goal. The answers to this question were already driving participants to think about what they would not like to do with their data. If there were already events that participants feared, we collected them on post-its.

4. In a next step, we specifically asked for unfavourable uses of data that participants saw when using the service in question, and also wrote them on post-its.
5. We then presented participants with several processing purposes that IT providers typically specify in regards to the aforementioned technologies. This list of purposes was informed by a hands-on search in privacy policies of websites, voice assistants and connected cars mobile application conducted by the researchers prior to the workshops and backed by one of the author's long years experience as a practising lawyer in data protection law. Confronting participants with the purpose specifications identified 'in the wild', we asked whether and how these purposes affected their previous 'risk' perception, in particular whether the purposes in question in any way limited the perceived potential uses of the data. In particular, we discussed the following purposes: 'Technical provision of the service', 'IT-security', 'service improvement', 'innovation of new services', 'making contact', 'personalisation of the service', 'advertising / marketing', 'transfer to third parties', and 'legal obligation'.
6. Finally, we asked users to jointly categorise the unfavourable uses of data collected and to discuss their understanding of the similarities between them. The explicit aim of this task was to categorise them as potential reference points for reformulated purposes of using data (by including some uses of the data more explicitly and excluding others).

3.1.3. *Data analysis*

All workshops were recorded and transcribed. Each document was analysed individually by two student researchers using the inductive coding process of thematic analysis. After the coding of the first workshop, the resulting codes were discussed and merged to a coherent code set. In case of conflict, the first and second authors were involved to discuss viable solutions for the existing codes.

In our coding, we concentrated on common patterns and arguments for an in-depth analysis when weighing up the identified phenomena. The analysis allowed us to follow the streams of argumentation of each participant in order to both preserve and enrich the unfavourable uses written down during the workshop. In particular, this enabled us to better understand the relationships between individual meanings of unfavourable data uses and their titles on the post-its.

3.2. *Findings of the user workshops: the wild list of 91 'risks'*

The analysis finally resulted in a list of 91 'risks', which we present below. In doing so, we highlight some interesting aspects that emerged from the analysis and that we consider to be highly relevant for answering the research question, also with regard to a later practical (effective) implementation. For reasons of space, we refrain from presenting the categories formed by the workshop participants and refer to the categories presented in the later chapter 3.4, which were merged with those from the experts.

3.2.1. *Similarly unfavourable perceived data uses for all three technologies*

In many cases, the unfavourable data uses collected in the various workshops were not clearly distinguishable, but often overlapped. In fact, as we conducted one workshop after the other, we observed a saturation of the collected unfavourable data uses. The more workshops we held, the less frequently the participants named new unfavourable data uses that had not been mentioned in previous workshops. Even though we do not claim that our list is a comprehensive list of unfavourable data uses, such an empirical saturation at least points into the direction of some generalizability.

Another interesting finding was that many unfavourable data uses, although they differ greatly in terms of technology, were repeated across the different workshops. Insofar, our workshops point towards the existence of a certain corpus of unfavourable data uses, which users may generally fear across different technologies regardless of the actual nature of data and context of collection. Such frequently mentioned instances include, for example, the fear of being tracked in general, or of being financially discriminated, or state actors using data to one's own disadvantage. The fact that the collected data uses are replicated across all workshops is the reason why we present all 91 collected unfavourable data uses (i.e. 'risks') in one consistent list.

This list was rather "wild", without any recognizable hierarchy or structure, similar to the list of risks as mentioned in recital 75 GDPR (Table 1).

3.2.2. *The perceived (severity and) likelihood of risks with respect to the purposes*

Although this list tends to be a saturated collection of unfavourable data uses for all three technologies, we observed differences in the likelihood and severity of occurrence of these individual 'risks' by workshop participants. We did not delve deeper into this question, but particularly the users' perception of the severity of such a 'risk' seemed to differ significantly, not least depending on their personal attitude. The perception of the likelihood that an unfavourable data use would actually come to effect also seemed individual, but more dependent on the specific use case and the technology as well as on the technological adeptness of the participants than on their personal attitude.

With particular respect to the likelihood of the respective 'risk', we asked how the processing purposes as specified by the technology providers played a role for the mitigation of participants' risk perception. Interestingly, in the further course of the workshops, it turned out that the aforementioned purposes typically stated by IT providers rarely lead to a meaningful restriction of the feared data uses. On the contrary, several workshop participants said that these purposes confirmed their fears to that extent that the unfavourable uses previously identified could and/or would indeed occur.

3.3. *Expert interviews to categorise the unfavourable data uses perceived by users*

As mentioned, the categories formed by the user groups were extremely wide-ranging and disparate. We therefore considered including a third perspective. For doing so, we consulted

Table 1 – Shows all unfavourable uses that participants came up with during the workshops, excluding duplicates that were removed by the authors.

adaption of behaviour	people who do not provide data will be financially discriminated	force or coercion to act in a certain way
analysis of daily living behaviour	people will be treated differently because of the nature of processing	hackability
being eavesdropped	personalised advertising	higher rates for car insurance
collecting personal data as a threat per se	personalised news / information bubble	higher rates for health insurance
collection of data about children	potential misuse of data	housing allocation or housing commitments is made more difficult or impeded
commercial exploitation of user data	powerlessness against technology / dissatisfaction using technology	identity theft (e.g. credit card fraud)
compensation claim if, for example, an ambulance is called due to a technical error	pressure to justify yourself / factual reversal of the burden of proof	inclusion in debtor register (including scoring)
contingency loss of the societal development	price discrimination	incorrectly personalised advertising
contravening societal interests	privacy as a good in itself	influencing someone's political opinion
cyberbullying	profiling (e.g. purchasing behaviour)	influencing the political opinion of many people
data about locations (and changes thereof)	publication (of e.g. photos)	information asymmetry (e.g. compared to website operators due to automation)
data subject is misused as guinea pig for product innovation	restriction of freedom	insights into one's personal life by humans or machines as a criterion in the privacy assessment
data trading	restriction of practising a religion	insights into the bedroom
data transfer to outside of the euu	revealing genetic /biometric data	law enforcement
denial of occupational disability insurance	revealing health data	less opportunities of societal retreat
denial of payment by liability insurance	revealing sexual behaviour / partner search	lock-ins when items are too personalised
difficulties in finding a job	revealing socio-economic status	lock-in effects (esp. user cannot move their data to another provider)
diffused anxiety	revealing trade union membership	losing control over user of data
disclosure of data covered by the obligation of professional secrecy	reversal of pseudonymisation	loss of control over collection of data
disclosure of data on ethnic origin	security breaches (availability, confidentiality, integrity)	loss of reputation
disclosure of payment transfer data	self censorship 'scissors in the head'	manipulation of behaviour (e.g. in one's choice of a transport system or how to move through a city)
disclosure of religious beliefs	social exclusion	manipulation of purchasing decisions
discrediting people by manipulating information about them	spoiling surprises	many data are accumulated and stored centrally
discrimination at the workplace	states as actors becomes malicious	mental health impairment
e-mail phishing	surveillance pressure (e.g. from household members)	monitoring of communication
enactment of a social credit system	the political system could flip over/collapse	monitoring of the own living space
enforcement of civil law claims	theft of intellectual property	monopolies / supremacy of private actors on economic market
exercising your job under surveillance	unconsciousness of data collection	non-profit / profit purposes as a criterion in the data protection risk assessment
false conclusion are drawn from personal data	uncovering lies	on the basis of the data, laws are passed which some citizens consider to be disadvantageous for themselves
feeling of nakedness	unsolicited contact requests (not only for advertising purposes)	waste of time / efforts needed from users
financial disadvantages		

with eleven external experts asking them to categorize the unfavourable data uses collected in the user workshops. Interestingly, the expert interviews revealed that the experts had similar difficulties in categorising the wild list of user-perceived 'risks' just as the users had. For reasons of space, we also refrain from presenting the categories formed by the experts and refer to the merged categories presented in the following chapter 3.4. Only the study design for the expert interviews will be presented briefly.

3.3.1. *Expert acquisition and sample*

The experts were recruited from the professional network of the authors. Participants include renowned legal experts in data protection risk assessments, members of data protection authorities, practicing data protection lawyers, and UXD researchers specialising in data visualisation and socio-informatics. The eleven participants came from Italy, France, Germany, Great Britain and Luxembourg.

3.3.2. *Interview design and analysis*

For the expert interviews, we wanted to understand how experts (including practitioners and academics) in the field of data protection would categorise the unfavourable uses perceived by users, in order to classify them in a meaningful way and form coherent units. For the introduction, we provided all participants with background information on the project and explained why we were seeking their views on the subject (i.e. to define categories as reference points to re-specify processing purposes in such a way as to include unfavourable uses of data and to exclude others more specifically, which may serve as a reference scheme for a textual refinement or even the design of icons). Before presenting the unfavourable uses, we translated them into English. We then conducted remote video interviews with the international experts. The experts' task was, to (1) read and understand all unfavourable uses provided, and (2) – analogous to the user workshops – sort all unfavourable uses into categories that would be suitable for re-specifying purposes more precisely and, e.g. to design appropriate privacy icons.

For the entire interview, we transcribed all unfavourable data uses into a trello board (trello.com) and into one trello list. Together with the interviewed expert, the interview generally began by asking the expert to go through all unfavourable uses to eliminate ambiguities and create a common understanding. In an analytically second step, we asked the interviewee to go through the list of unfavourable data uses, and create and name new lists for categorising unfavourable uses. We asked the participants to strive for a clear classification of all unfavourable uses. However, if participants had difficulties, we allowed them to make copies of unfavourable uses and categorise them under more than one category. During this process, we used screen sharing, and asked the participants to 'think aloud' to follow the respondents' considerations.[85] As good practice, we have occasionally asked respondents to explain their reasons for categorising and naming a particular category.

[85] T. Boren, J. Ramey, 'Thinking Aloud: Reconciling Theory and Practice' (2000) 43 IEEE 3, pp. 261-278.

The interviews lasted about 90 min each and were recorded, in most cases, on video for later analysis. For the expert interviews, our analysis focused on the categories and hierarchies provided by the interviewees. Here, the first and second author discussed meanings and interpretations collaboratively after the interview in order to recapitulate and conserve the understanding of the lists drawn up by the experts.

3.4. *The merged data use ('risk') categories as a reference scheme for effective purpose specification*

Despite the variety of approaches to categorise the list of 91 'risks', several common patterns could be identified. The following sections describe the different categories that resulted from the categorisation process.

Socio-technical environmental variables were understood to have no direct impact on the data subjects, but rather formed a socio-technical condition under which data processing would take place. These variables included circumstances that one expert called – having the data processing principles under Article 5 GDPR in mind – 'Risks to existing legal principles/data processing procedural principle/due process concerns'. Another non-legal expert grouped them under the term 'security', under which we classified other 'risk' classes such as 'inaccurate data', 'data integrity', or 'malfunctioning', but also 'identity theft'. The reason for classifying identity theft under this category was that users did not perceive the 'risks' associated with such theft as an immediate impact on their lives, but rather feared the potential consequences of such an identity theft in the future, e.g. loss of money (when the thief uses the identity to withdraw money from the user's bank account) or damage to reputation (when the thief uses the user's identity to make defamatory public statements on behalf of the user). Other categories referred to the perceived control (in the sense of 'loss of control') or transparency (in the sense of 'lack of transparency') of processing or trustworthiness of actors involved. Factors for what we call trustworthiness are, for example, whether collected personal data are used for economic purposes (i.e. are commercialised), or whether the state or private entities process the data.

Potentially more impactful is the category of **privacy intrusion,** an umbrella term containing a large number of similarly named categories including 'privacy', 'private feeling', 'private / state surveillance', 'unwanted intrusion' 'unwanted disclosure' etc. We considered this category to be potentially more impactful than the category of socio-technical variables mentioned above because workshop participants described the 'risks' covered by the 'privacy-intrusion' category as having a direct impact on their private lives, regardless of how the disclosed data is ultimately used. Thus, users considered the intrusion to have an impact on their lives simply because somebody else is receiving information about their private life. Some users also distinguished between different privacy-related issues. For example, with respect to the 'human in the loop' debate surrounding voice assistants, one workshop participant said he had not considered the pure processing of private information by an algorithm as an intrusion into his privacy (because such a privacy intrusion apparently requires, in

his opinion, a human who gets the private information). On the other hand, another participant said that the more people have access to such information, the more conspicuously their privacy would be concerned. The generation of further knowledge about a data subject by analysing the accessed information, e.g. through profiling, was also clearly considered as an additional insight into their private life. Some experts even sorted profiling as a separate category. However, since profiling means collecting information about someone's private life and generating even more insights through inferences, we considered this as a sub-category of privacy.

In contrast, in further categories concerning how data use may unfavourably impact data subjects, the use of algorithms was usually considered more problematic than in the privacy intrusion category. In this regard, we observed a couple of smaller categories referring to different (immediate) effects on data subjects. For example, **behavioural manipulation** was a frequently mentioned risk category easily distinguishable from other categories in the sense that it consists of attempts to influence behaviour or control activity. The use of nudging techniques also played a major role in this regard. Accordingly, the restriction of freedom through such manipulation techniques was also assigned under the umbrella term of behavioural manipulation.

Negative effects on social status and/or relationships formed a fourth cluster, which could be divided into the impact of social relations of an affected data subject. Perceived negative effects can occur on the general level, in the context of public relations (including politics), in the professional sphere, within family life, as well as within friendships. The general level includes situations such as 'pressure to self-exposure' or 'loss of reputation', 'risks' that occur in any social context. The subcategories refer to more specific contexts, including the political context in terms of public relations, involving concerns about facing political persecution based on the information collected. Another subcategory refers to professional relationships (e.g. where 'risks' perceived by users may make their 'professional life more difficult').

The 'risks' of **health impairment** were also apparent in our empirical work, and experts in particular formed separate categories for health-related 'risks'. This umbrella category deals with unfavourable uses of data, e.g. in the area of health insurance, as well as the impact of data collection on personal health. The merging of expert and participant categories also reveals a clear category related to **material harm**. One instance representative of other 'risks' subsumed within material harm is the self-explanatory 'risk', 'Could cause you to lose money'.

Workshop participants and experts also saw potentially unfavourable data uses in the area of **discrimination**, covering subcategories like 'discrimination', 'social disadvantages', or 'impairment of participation'. In contrast, unfair and unfavourable treatment of data subjects is also associated with the risk category **law enforcement**. This category covers situations where data could be used for legal persecution, undermining data subjects in legal proceedings: for instance, in their ability to present their version of events (in practice, this could result in the reversal of the burden of proof).

Generally, both experts and participants envisioned **societal risks** as categories of unfavourable data uses. These often focussed on undesirable (even unintentional) ways of shaping collective life by data collection and use: for example by supporting undesirable businesses or business practices. The experts also added categories called '**abstract meta risks**', '**no data protection risks**' and '**other**' to indicate that they did not consider it appropriate to mention those risks in a direct data protection context.

4. Implications for the legal discussion

The results of our empirical-qualitative research suggest several implications for the legal debate. While our study may not have directly answered our research question of how to effectively specify processing purposes, we identified a set of 'unfavourable data use' ('risk') categories that can be used as a reference scheme to more clearly formulate processing purposes either including or excluding such uses ('risks') in future. An orientation towards these categories as a major resource for speaking the language of the user can significantly help to be more transparent in communicating use of data and finally avoiding unexpected, inappropriate or otherwise objectionable data use. In particular, our research process showed five main implications for the interpretation of the law.

4.1. *The current implementation of purpose limitation fails*

The most apparent result is the failure of the current implementation of the principle of purpose limitation. More precisely, the current implementation formulating purposes as 'technical provision of the service', 'IT security', 'service improvement', 'innovation of new services', 'making contact', 'personalisation of the service', 'advertising / marketing', 'transfer to third parties', and 'legal obligation', at least without further information, fails the first component of the purpose limitation principle: to specify processing purposes in such a specific manner by which data subjects can assess whether they find the data use at hand 'unexpected, inappropriate or otherwise objectionable'.[86] The reason for this is that almost all participants answered to the question of whether these purpose statements would in any way restrict the previously collected 'unfavorable data uses' in the negative. On the contrary, participants concluded that most of the purposes rather confirmed their fears. This result does not mean that the principle of purpose limitation is inherently faulty (this was the instantaneous reaction of some experts on our preliminary findings). Rather, the empirical results suggest that the way *in which* purposes are *currently* specified do not meet the regulatory objective. Moreover, our empirical findings suggest that the reason for this failure is not an overly ambitious regulatory objective that is unattainable in praxis. In fact, not all of the currently (typically) specified purposes seem to fail in the same way.

From the perspective of the user workshop participants, there are less problematic purposes such as the 'technical pro-

[86] See our research question in chapter 2.1.3 referring to Art. 29 Data Protection Working Party, 'Opinion 03/2013 on Purpose Limitation' (WP 203, 02 April 2013), p.11.

vision of the service' and 'IT security'. However, these users might even underestimate the unfavourable uses that could potentially result from processing their personal data for IT security purposes. For example, in the instance of a DDoS attack on a website, users could be suspected of committing the attack if their home routers (using their IP address at the moment of attack) was taken over by a botnet without the knowledge of the user. In such a case, the storage of dynamic IP addresses for eventual IT forensics in relation to a denial of service attack could lead to users being suspected of such an attack while their home routers (which were using this IP address at the time) were taken over by a botnet without the knowledge of that user. Thus, such stored IP addresses can be used against data subjects in the course of a cybercrime proceeding. Such data subjects may therefore want to assess the correctness of that data and challenge such findings. This ultimately indicates a need for data subjects to assess the appropriateness of data use in respect to their right to a fair trial.

Participants were less clear, however, when it came to the purposes of 'service improvement', 'personalization of the service', 'advertising / marketing' and 'legal obligation'. Participants were unsure to what extent each of the aforementioned purposes could be defined. Specifically, they were unable to independently assess whether personalised advertising could lead to price discrimination, or whether an improvement in the service meant an improvement in their favour or rather in the favour of the IT provider, an effect which could be detrimental to users. For example, as part of the service improvement, a manufacturer (or provider) of connected cars could start offering a personalised insurance based on collected driving data, leading to a poor rating. These concerns have become even greater in the context of 'innovation of new services'. Most strikingly, for the purpose of 'transfer to third parties', participants were unable to understand what third parties would do with data they received, not least because the third parties remained unnamed. In this regard, participants mentioned that such a purpose would be a 'blank cheque' for the use of data in any possible way.

In this respect, two aspects need to be clarified from a legal perspective. Firstly, the mere fact that many data controllers specify these purposes in the aforementioned manner does not mean that this approach is legally correct. As the EDPB has pointed out, 'a failure to state, or accurately state the purpose or purposes for processing does not mean that the data controller can process personal data for any and all purposes at its discretion'.[87] Rather, the actual purpose must be reconstructed according to the facts of the case – and an excessively broad purpose may well lead to a ban on the processing itself.[88] For example, if a specified purpose in a data subject's consent form does not sufficiently indicate a certain data use that data subjects perceive relevant, this data use cannot be based on the consent.[89] Secondly, the findings of our empirical research suggest that there is a common understanding of 'risks' amongst users to which a data controller may refer to in order to more clearly specify certain uses of their data (while excluding others). This leads us to the second implication for the legal debate.

4.2. *The common perception of 'risks' across different technologies (and its taxonomies)*

The second interesting implication for the legal debate is the fact that there may be a common corpus of unfavourable data uses across different technologies as perceived by users. The unfavourable data uses that the participants in our user workshops offered not only show a tendency toward qualitative saturation, but also decidedly replicate across all three technologies (i.e. using websites, voice assistants, and connected cars). This finding is important because it points to a potential standardisation of processing purposes through the use of the proposed scheme of categories (or a similarly generated one), a standardisation which is more or less independent from the technology in question.[90] For example, in a next step, we could re-specify the purpose of 'IT Security' more precisely: for instance, one can indicate that the collected IP addresses of data subjects might be used in legal proceedings in relation to cyber attacks against them (e.g. if their home router were to be hijacked by a botnet to carry out denial of service-attacks). Such a re-specified processing purpose could be used across different technologies (using appropriate examples) and could consequently be represented through a common graphic icon according to Art. 12 sect. 7 GDPR.

4.3. *The interplay of several 'risks' with respect to one technology (more precisely, one processing operation)*

Another relevant implication is that one and the same processing operation can lead to simultaneous exposure to several 'risks'. This implication results from the observation of how difficult it was to draw clear lines between unfavourable data uses through distinct categories. One expert made an explicit comment in this regard when sorting the user-perceived data uses and noted their impact on both the individual and societal level. According to this expert, it was impossible to exclusively sort certain data uses under the category of individual or societal impact, since the use of personal data to a certain extent always presents a level of influence on the individual and societal level. The same problem became apparent with respect to further data uses. For example, the workshop participants mentioned the 'risks' of 'loss of privacy / private data'. It is difficult (if not impossible) to say whether this phenomenon should be sorted under the category of 'impact on privacy' (i.e. 'loss of privacy') or be considered a loss of control over personal data (i.e. 'loss of private data'), which would

[87] Art. 29 Data Protection Working Party, 'Opinion 03/2013 on Purpose Limitation' (WP 203, 02 April 2013), p. 19.

[88] Ibid. p. 19 footnote. 50.

[89] M. von Grafenstein, '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)', going to be published in EDPL 03/2021.

[90] M. von Grafenstein, '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)' (2011) going to be published in 5 EDPL 02/2021, referring to B.-J. Koops, 'The (in) Flexibility of Techno-Regulation and the Case of Purpose-Binding' (2011) 5 Legisprudence 2, p. 171.

be considered as a socio-technical condition (e.g. influencing the *impact* on the user's privacy). Another example is the phenomenon of 'chilling effects' (i.e. users cease or omit a certain action because they are afraid that somebody else will find out and do something about it). This 'risk' could be classified under the category of privacy, as privacy is supposed to protect against chilling effects. Likewise, such a 'risk' could also be categorised under 'limitations of freedoms', which are the result of these chilling effects. Furthermore, the risk of using 'sensitive data' could be interpreted as being directly linked to the privacy of an individual (since these data points reveal sensitive aspects about the individual's private life) or as a socio-technical factor that increases the likelihood and severity of the risk (for instance, that this information is used against them, leading to negative consequences for the data subject). Similarly, unfavourable use of data to increase the price of a health insurance policy could be placed in a health-related category or in the category of financial loss. Likewise, the 'risk' of 'economic influence' could also be placed in the category of financial loss or manipulation of behaviour. These examples demonstrate that unfavourable data uses cannot always be clearly assigned to one category or another. However, as these categories are not necessarily mutually exclusive, this problem does not have to have great impact in practice. In principle, processing purposes can be re-specified in a way that indicates a 'risk' not only in one or the other category, but in both or even several categories.

Lastly, it is evident that the presented categories do not present a clear hierarchy. For example, the privacy category implies a set of more specific subset categories such as privacy at the home, privacy in public, or privacy when using communications. Similarly, the category of social relationships includes subcategories like relationships amongst family and friends, at the workplace, or even in the public relations context. In fact, most of the presented categories could be further merged, for example, into three main categories: privacy intrusion, impairment of freedoms (including discrimination), and material harm. However, the precise taxonomy does not have to be determined here. Rather, it is reasonable to assume that the precise taxonomy again depends on what the data subjects find most effective in a specific situation. This opens up a more specific research question than has been addressed in the current stage of research.

4.4. *Mapping the data use ('risk') categories to the data subject's fundamental rights*

Another interesting observation is the similarity of the categories to the data subjects' fundamental rights. As mentioned previously, the proposed categorisation offers only one example which could be adopted to improve the principle of purpose limitation from a users' point of view. Other categorisations – both new or pre-existing – are equally conceivable: as mentioned previously, one specifically suitable alternative are the data subjects' fundamental rights. In fact, many of the unfavourable data uses perceived by the users that we identified corresponds to a fundamental right. For example, the category of 'privacy' most likely corresponds to the right to private life under Art. 7 ECFR. The category of 'manipulation' could have its equivalent in the right to data protection as stated in Art. 8 ECFR, and eventually in respect to other rights such as consumer protection according to Art. 38 ECFR. The social status of data subjects or their social relationships in general (e.g. affected by loss of reputation or pressure of self-exposure) may again correspond to their right to private life (Art. 7 ECFR), while specific social contexts may correspond to the data subjects' specific right to privacy of family life (Art. 7 ECFR), their right to engage in work (Art. 15 ECFR), their citizens' rights according to Art. 39 and 40 ECFR, or their freedom to assembly under Art. 12 ECFR. 'Discrimination' may correspond to the equality rights under Art. 20 et seq. ECFR, and the risk of legal proceedings to the data subjects' justice rights according to Art. 47 ECFR. 'Health impairment' can correspond to the right to respect for mental integrity (Art. 3 ECFR) or the right of access to preventive health care (Art. 35 ECFR). 'Material harm' may find its equivalent in the right to property under Art. 17 ECFR. Last but not least, the 'societal risks' may correspond to other constitutional guarantees such as solidarity according to Art. 27 ECFR (in addition to consumer protection already mentioned).

As a result, the inductive-empirical approach (applied in this paper) and the deductive approach (which refers to a legal framework such as the data subject's fundamental rights) not only correspond more or less, but may rather complement each other. On one hand, legal reasoning is methodically well-equipped to carve out the substance of values (which have been established through fundamental rights and further defined by ordinary law) on a general basis. The canon of all fundamental rights also demonstrates that the list of user-perceived risks should not be misunderstood to be comprehensive nor exclusive. In fact, there can be numerous fundamental rights that provide a substantial guarantee that data protection law aims to protect, but whose significance the participants in our user workshops have not (yet) acknowledged. Thus, the normative value and importance of such a fundamental right is not reduced because the workshop participants have not thought about it.[91] On the other hand, legal experts may refer to empirically assessed expectations of data subjects in order to concretise the normative substance of certain fundamental rights (in particular the privacy rights in Art. 7 ECFR). On this basis, a user-centred process can facilitate understanding how such rights might be effectively implemented in a specific situation. Indeed, combining both disciplines creates a field of tension: while UXD research takes users' thinking and sayings into the centre of their research, and therefore assigns high value to user opinion, from a legal perspective, a user opinion is just one interesting (but important) factor to be considered in legal reasoning. However, this field of tension can be solved in practice by consulting both perspectives in order to clarify misconceptions and balance contradictory views: here, mediation between both sources is critical. The aim here is that such an interdisciplinary approach leads to an implementation of the law that is, as required, more effective.

[91] See already above at point 1.3, referring to A. Voßkuhle, 'Grundlagen des Verwaltungsrechts', § 1 Neue Verwaltungsrechtswissenschaft (2012, 2nd edn., C.H.Beck) pp. 22-28.

4.5. *Aligning the user perspective with legal terminology (esp. purposes, threats, risks, consequences, impact)*

Last but not least, another interesting result for the interpretation of the law is that users show a clear tendency to consider the 'impact' or 'consequences' of the data processing on their lives as most relevant in the course of evolving data processing 'risks'. How important this result is for the discussion on data protection law becomes clear on closer examination of the diverse and not yet fully clarified terminology in the GDPR. The Data Protection Directive, which preceded the GDPR, still mainly focused on the purpose of the processing in order to capture the relevance of the data processing for the data subject from a legal viewpoint, and otherwise remained conceptually sparing with semantically similar terms such as 'risk' and 'consequences'.[92] The view that purposes should show the consequences for the person concerned could thus be defended relatively easily without getting entangled in a maze of semantically similar words which, according to the logic of legal reasoning, must nevertheless be distinguished from each other. This has changed with the GDPR. In addition to the processing purpose, the GDPR now explicitly introduces other essential regulatory concepts. These include above all the risk-based approach (with its focus on the 'risks to fundamental rights') and the data protection impact assessment (which additionally focuses on the 'impact on the protection of personal data' and, methodologically, distinguishes both terms from so-called 'threats'), but also the now extended protection against automated decisions (according to which the controller must inform the data subject of the consequences of such decision-making systems). This raises the question of how all these terms are conceptually interlinked or, vice versa, to be distinguished from each other.[93]

These questions can have a considerable influence on the concrete application of the GDPR in a specific case. As described in the introduction, it is controversial in the legal discussion, for example, whether the fundamental rights of the data subjects should be used directly as a reference point for the risk assessment and thus also for the effective implementation of the protection measures. From the point of view of the data subjects, on the other hand, the relevant reference point seems to be clear: for them, it is primarily a question of whether the data are used in a way that invades their privacy, restricts their exercise of freedom, discriminates against them and/or brings them material or other disadvantages. As shown in the previous chapter, these user-perceived 'risks' correspond to a large extent to their privacy, freedom and participation rights. From the user perspective, it is therefore obvious that the controller must implement its protection measures in such a way protecting these rights directly.

Further, it could be argued from a theoretical-legal point of view that the processor does not have to specify its processing purposes in accordance with Art. 5 sect. 1 lit. b GDPR in a way showing the 'consequences', 'risks', 'impact' or 'threats' for the data subject, because all these terms are already covered by other provisions. From the point of view of the data subjects, however, the question is clear: from their point of view, it is precisely this level that is at stake, i.e. the risks to their fundamental rights. In our opinion, the legal-conceptionally right idea of distinguishing as sharply as possible between different semantically similar terms within a law should therefore not be overstretched. Rather, against the background of the extremely complex legislation process,[94] one may legitimately ask whether the legislator intended to link different legal consequences to the different terms at all or whether they do not rather represent different entry points into one and the same risk assessment (albeit with different analytical focuses).[95] At least, in contrast to the ambiguities of the legal language in the GDPR, the issue at hand appears much clearer to data subjects themselves.

5. Outlook: setting the new 'state of the art'?

As shown in the introduction, the idea of making data protection law more effective by requiring the regulation addressee to implement the legal norms into the processing design itself is rather old. In fact, this concept is nearly as old as the data protection and privacy debate. However, there is a more recent shift towards establishing the data protection by design approach within law, and the implications of such an approach are not yet fully understood in the data protection debate. As mandated by Article 25 GDPR, the regulation addressee is required not only to implement the legal norms into the processing design, but to do so in an effective manner. By explicitly declaring the effectiveness of the protection measures to be the legally required result, the legislator raises the question of which methods can be used to test and ensure such effectiveness. In fact, extending the legal conformity assessment to the real effects of the required measures opens this assessment to (non-legal) methodologies that are specialised for assessing such empirical facts. This does not mean that lawyers must directly incorporate non-legal methodologies and findings into the legal interpretation of Art. 25 GDPR or even apply these empirical methods on their own. Rather, the findings of non-legal disciplines are usually taken into account as a factor in the interpretation of the law. The consequences of the inclusion of this factor may pose significant in terms of legal practice: Interpreters of the law (e.g. a

[92] The Data Protection Directive 95/46/EC uses the term 'purpose' 59 times, the term 'risk' eight times, and the term 'consequences' once, <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31995L0046&from=EN> accessed 18th January 2022.

[93] In contrast, the GDPR uses the term 'purpose' 261 times, the term 'risk' 75 times, the term 'impact' 31 times, the term 'consequences' 14 times, and the term 'threats' 10 times, <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=DE> accessed 18th January 2022.

[94] One indication of this may be the fact that almost 4,000 amendments were tabled in the legislative process, which had to be processed accordingly by the editors of the legislative text, <https://edri.org/our-work/data-protection-series-issue-sheets/> accessed 18th January 2022.

[95] M. von Grafenstein, '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)', going to be published in EDPL 03/2021.

data protection authority or a legal court) cannot simply ignore methodologically-assured findings of the external discipline, since these practices and methodologies describe the factual situation on which the interpretation of the law is based. In the best case, lawyers collaborate with representatives of these other disciplines (e.g. informatics, UX designers) aligning their respective concepts and methodologies to make the implementation of the law more effective.

In fact, this interdisciplinary opening in Art. 25 GDPR fits into a larger development in the regulation discourse. Under the label of evidence-based policy-making, a discussion has been going on for quite some time now about how to increase the rationalisation of law by referring to non-legal disciplines. There are possible pitfalls to this approach, including increased complexity when considering the effects of regulation instruments in legal reasoning. In the course of this contribution, we kept returning to these challenges and offered possible solutions. In particular, we have demonstrated how both disciplines of legal and UXD research can adopt elements from one another. More interestingly, interdisciplinary work between the two fields can also have the consequence of influencing the respective practices and conceptual reasoning contained within the fields themselves. Ultimately, this may prove beneficial for both fields in achieving their own field-specific objectives. These insights may be transferable to other data protection principles and rules, whose effectiveness assessment requires input from other non-legal methodologies. In our current example of purpose specification, we demonstrated how an interdisciplinary approach can help to specify processing purposes in a more effective manner. This would allow for the formulation of purposes in a way that more explicitly 'prevent[s] the use of individuals' personal data in a way (or for further purposes) that they might find unexpected, inappropriate or otherwise objectionable'.[96]

To this aim, we have presented a qualitative approach and first findings about which uses of personal data laypersons (i.e. data subjects) consider relevant. These categories of unfavourable data uses may serve as a reference scheme for data controllers when specifying their processing purposes in future practice. Using these categories as a reference scheme for purpose specification could ensure that the specified purposes clearly indicate the use of personal data that data subjects find relevant. Consequently, this offers data subjects transparency and reliable expectations on how their data will be used. Thus, the proposed categories of relevant data uses may enable controllers to meet the regulatory objective of the principle of purpose limitation and to effectively implement the principle according to Art. 5 sect. 1 lit. b and Art. 25 sect. 1 GDPR. Since these data uses fall under the scope of the fundamental rights of data subjects, these rights can serve as a direct reference to further adjust the protection measures.

In future research steps, we may focus on different elements by continuing to apply a user-centred, multi-stakeholder approach. First, we may begin by refining typical processing purposes on a textual level by conducting workshops together with data controllers and data protection authorities. Together with these stakeholders, we can clarify the processing purposes by referring to our category scheme. Secondly, in collaboration with laypersons and professional graphic designers, we may design icons representing our proposed categories of data uses. Lastly, it is equally possible to focus on testing the effectiveness of current implementations of the GDPR: for instance, the data subject's consent. For example, there has been plenty of research showing how currently dominating implementations of cookie banners fail to properly inform users and deceptive design patterns, which are now also acknowledged by law. In this contribution, we have shown how to proceed methodically to test such effectiveness and comply with Art. 25 GDPR.

Thus far, our research results have been based on qualitative methods. While generalisable, they are not representative (of a general population such as of Germany or even the EU). Still, qualitative methods can be evidence of effectiveness.[97] The highly interdisciplinary nature of UXD, however, also provides roads to make these empirical results even stronger, for example via triangulation with quantitative methods. In any case, the legal implications of our current and future results should not be underestimated. The legal discussion, or in a more practical context, data protection authorities and legal courts may come to the conclusion that certain implementations of legal transparency and control requirements based on our scheme provide for more effective protection against certain risks than other ones. If this is the case, these more effective implementations constitute the new state of the art according to Art. 25 Section 1 GDPR. Controllers that do not implement these requirements in an equally effective way run the risk of noncompliance. Of course, data controllers may always provide counterevidence: for instance, to argue that their specified processing purposes (or their consent mechanism nudging the data subjects into one direction or another) are equally or even more effective. In such a case, however, data controllers must be careful about the methodology they use to ensure that their results are scientifically rigorous and reliable. They might use the same or a similar methodology as presented in this contribution.

Declaration of Competing Interest

The following research results are part of the interdisciplinary research project "Privacy Icons", an ongoing project conducted at the chair Digital Self-Determination at the Berlin University of the Arts. The chair is part of the Einstein Centre Digital Future, a public-private partnership that is partially funded by the Berlin government and partially sponsored by private entities, e.g. the Deutsche Kredit Bank. At the point of publishing this article, there has been no involvement of the above-mentioned sponsors (or any other additional sponsors) in the "Privacy Icons" research project. A future involvement is planned for a later stage in order to fulfil the requirements of data controllers with respect to purpose specification.

Data Availability

Data will be made available on request.

[96] Art. 29 Data Protection Working Party, 'Opinion 03/2013 on Purpose Limitation' (WP 203, 02 April 2013), p. 11.

[97] EDPB, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default' (adopted on 20 October 2020), version 2.0.