Bringing Context back into privacy regulation and beyond

About limitation on purpose as an (old) response to (new) data challenges

Karoline Krenn

Introduction

At the core of early privacy debates were state records, corporate records or survey data. The advancement of information technologies extended the availability of data. New technologies mediate many aspects of modern life and, thereby, enable data to be circulated. They provide access to very different types of data from very different sources. Along with that goes a strong power asymmetry between the individual users and the organizations involved in the industrial processing of data. The digital economy builds on access to individual data as fuel for its derivative operations, and government authorities respond with different directives to balance these asymmetries and protect the rights of citizens. The regulation of privacy reflects both a national and a supranational protective approach towards information infrastructure.

In response to the challenges in the digital age, public and private bodies introduced a set of privacy principles aimed at protecting individual rights. In 1980, the OECD formed the first internationally agreed-upon statement of core privacy protection principles, which were taken up and developed further by many governments and organizations (OECD 2011). By 2011, the International Organization for Standardization (ISO) had also published a privacy framework. The European Union Directive 95/46/EC from 1995 was the first serious attempt to implement privacy principles in a supranational regulatory framework. The first European Data Protection Law, the European General Data Protection Regulation 2016/679 (GDPR),[1] which has been in force since May 2018, is the EU's first comprehensive response to the challenges to privacy. It makes the common regulatory framework directly binding and mandatory and consequently more coherent for the member states.[2]

The core of all sets of privacy principles is limiting the collection, processing and storage of personal data to lawful and fair practices (OECD 2011). To those principles belong the specification of purposes for which personal data is collected ("specification of purpose principle") and the limitation of use to these purposes ("limitation of use principle"). The former states that the purpose of the information must be stated explicitly and the latter stipulates that data cannot be used for purposes other than those specified, except with informed consent or by the authority of law. The "data quality principle" concerns the accuracy and completeness of data. The "security and safeguard principle" points to the safety of data against unauthorized use. The "openness principle" requires transparency about developments, practices and policies with respect to personal data. The "individual participation principle" demands individual access to and the ability to challenge one's own data. And finally, the "accountability principle" expresses the operator's responsibility to comply with these principles. To a large extent these principles overlap between frameworks, although their semantics and combination vary. With regard to their structure there has been little attempt so far to address how these principles relate to one another and what principle should be applied first. Auditing methods for privacy protecting systems do prioritize specification of purpose, but without much explanation (Makri and Lambrinoudakis 2015).

In this article I will focus on the limitation on purpose principle (LoP). In the GDPR, the "principle

Karoline Krenn is sociologist and researcher at the Fraunhofer Institute of Open Communication Systems (FOKUS) in Berlin. She is currently working on challenges of digital societies such as data sovereignty and privacy, the socio-economic dimensions of socio-technical design, and data methodology. She is the editor of "Markets and Classifications. Categorizations and Valuations as Social Processes Structuring Markets", a Special Issue of *Historical Social Research* 42(1) (2017). *karoline.krenn@fokus.fraunhofer.de*

of purpose limitation" unifies two other principles: the "specification of purpose principle" and the "limitation of use principle". In the following I will neither give a detailed account of how LoP operates in practice nor how it interacts with other principles. What I will do is to argue why LoP is particularly apt to respond to privacy challenges and what we can learn from the German debate about the impact of LoP. The purpose of data is an overall defining criterion contained within several principles such as specification of purpose, collection of data, as well as limitation of use. LoP is frequently singled out as an especially important principle, although it presents challenges in practice (Bygrave 2014). There are three particular reasons why I focus on limitation on purpose. First, it is particularly apt to define information domains avoiding the public-private distinction, which characterizes many privacy debates (Pohle 2015). This is particularly relevant with regard to online data for which it is often hard to tell if it is private, public, or both at the same time. Second, purpose refers to the context of data generation, which has relevant implications for the interpretation of what we can learn from data. This brings me to my third reason. De-contextualization generates a specific uneasiness because of the widespread use of data for (automated) decision-making by government agencies and businesses. A strong skepticism towards decision processes based on selected pieces of decontextualized information ("the data shadow") already characterized the European Directive of 1995 (Mendoza and Bygrave 2017). The partiality and shallowness of such decisions were considered as dehumanizing and making fully automated decisions was forbidden.

LoP has been implemented in German regulation since 1971 and has shaped the European debate since then (Pohle 2018). Reaching back to the 1970s, I describe the "context turn" in the German debate and how it influenced LoP. The debate shows that bringing context back in, first, shapes the understanding of privacy, and, second, provides a methodological criterion for data accuracy. This focus is also reflected in the literature. Context has regained prominence as a theoretical framework for privacy during the last decade (Nissenbaum 2009), although with distinction from the purpose approach. Nissenbaum critizes LoP for having "only indexical meaning" (Nissenbaum 2015, 291), lacking substantive criteria to specify purpose and leaving the protection to the controller's discretion. Recent literature addresses this critique and explores a framework for LoP from a legal viewpoint (Grafenstein 2018).

This article will proceed as follows. After a discussion of the challenges in a digital society and how LoP responds to them, I will explain the stipulation of LoP within the GDPR. I then turn to the German privacy discourse and regulation of the 1970s to show that data context was already perceived as relevant at this stage of information technology. In the section on digital mass data I examine the methodological limitations of de-contextualization. Bringing these two debates together opens up an additional perspective on the forms LoP can take and what constitutes its strength to control processes of information flow. The final section discusses the limits of the consent requirement for derogation from LoP based on recent cases of data repurposing.

Digital challenges

The challenges of the digital transformation of society have recently received increased public attention. Despite promises to facilitate social participation and advance transparency, societies are witnessing increasing inequalities. This has stirred debates identifying digitization as an actual driver of social inequality and rising social polarization. Initially the focus was on the labor market, arguing that a technology and skill driven economy is favoring capital and a minority of highly skilled individuals (Acemoglu 2002; Brynjolfsson, McAfee, and Spence 2014). The growth of tracking and surveillance technologies, sensor networks and compiled databases made information exchange a subject matter for critical debate. The volume of data generated and circulated is reaching the petabyte-scale, fueling various dynamics. These technologies themselves generate social differentiation (Gandy 2009, Fourcade and Healy 2013, Pasqual 2015, Poon 2016). New instruments for monitoring, sorting and profiling affect people on multiple dimensions: They segment markets and increase social inequality. They force cultural and political conformity, encroach on the moral autonomy of the individual, and threaten democratic principles.

Data is used for profiling and microtargeting in various domains. Microtargeting has long been a widely applied strategy in marketing. However, the digital infrastructure provided by online platforms and mobile applications has opened up new opportunities to record behavioral traces and differentiate consumers. It has created permanent surveillance (Zuboff 2019; Sevignani 2017). In addition to familiar market records from electronic payment data, customer profiles or loyalty programs, recent studies illustrate the extent of the tracking of basically every digital activity (or lack thereof) (Christl and Spiekermann 2016). Online participation and communication are turned into a huge profiling database. Clicks, likes, swipes, web searches, flows of communication

and geo-locations are recorded and compiled. Data are aggregated into categories, often designed as behaviorally defined risk groups, to increase efficiency and to predict outcomes, promising greater profits for commerce and protection against high-risk customers. The tech industry is driven by the prospect of monetizing data. However, business models that rely on data harvesting are most often opaque, and the flows of data are non-transparent to the average internet user.[3]

These efforts to detect patterns have a downside. Statistical profiling of online data is not a neutral tool but carries biases. An experimental study using a simulation tool that measured the use of information by web advertising algorithms and by personalized ad settings showed that, if information on the gender of users in search of a job was included, males were significantly more likely to receive ads encouraging coaching services for high-paying jobs than females (Datta, Tschantz, and Datta 2015). This is just one example of how digital profiling might systematically discriminate. Moreover, algorithmic sorting repeats existing patterns. Recommender systems expose digital media users to more of the same content and reduce new encounters. Thereby, sorting affects social connections and cultural experiences. This points to the cultural challenge of these new technologies.

The social effects of algorithmic sorting and profiling depend on the domain of application. It generates various kinds of classification situations (Fourcade and Healy 2013). Personalized ads and special offers can be annoying and price discrimination may contradict ideas of fairness. But there is also clear informational harm and inequality (Hoven 2001). The inclusion or exclusion from chances of market participation such as particular job or housing offers severely impact life chances of individuals. It reinforces existing inequalities between groups. And these classification situations generate inequalities on novel dimensions specific for digital technologies. These risks grow when data is exchanged between the private and public sector. China is an interesting illustration of a blurred interplay of those two. Its Social Credit System illustrates the extent to which such a punishment-reward-system can be escalated (Liang et al. 2017). There, recorded non-conformity to rather strict social norms and beliefs lead to exclusion from basic public goods such as education or transport.

There is also a political challenge. Unbalanced access to information and potential manipulation also conflict with the self-understanding and value system of a democratic society as they defy individual rights. Societies have to deal with new polarizations. This is quite obvious in the political domain where tailoring information encroaches on the autonomy of the individual and threatens civil liberties and democratic principles (Hoven 2001). Microtargeting of potential voters, echo chambers in social media news feeds and filter bubbles pose major risks for the political opinion building processes. The Brexit vote and the US elections in 2016 are two of the best examples.

These challenges intensify with the proliferation of intelligent homes and urban spaces equipped with sensors, and with administrative processes becoming more and more tied to complex data. Consequently, they require a continued debate on "good" and "bad" data usage. Particularly, and aggravated by data driven automated decision-making, patterns of inclusion and exclusion are likely to be even more shaped by socio-technical arrangements in future digital societies.

As diverse as these challenges are, they are intensified by an unregulated repurposing of data. Information technology makes it easy to access and to combine different information sources and to compile data collected for different purposes and from many different contexts. This raises a general problem in a data-driven society: How to handle the multiple future usages of data whose use is not restricted at the moment of collection? This problem is made worse by the power asymmetry between organizations servicing the digital infrastructure and the individuals providing data.

The challenges in digital societies point to fundamental underlying conflicts of interest and values. The domain of information exchange is just one, albeit important, stage for potential conflict. Regulation responds to these challenges by limiting the processing of data. LoP is effective in particular because it regulates repurposing. The purpose frame allows linking the specification of purpose at the time of collection with those of further processing. It connects different contexts of usage (Grafenstein 2018) and provides a criterion for appropriate data use (Pohle 2015). Moreover, LoP not only addresses civil ideals such as informational self-determination, but, due to its link to data contexts, it also responds substantially to the main social challenges: the harms to the individual through the mixing of information from different social contexts. This conflation is a major gateway for the spread of disadvantage from one social domain to others, as has been shown for the off-label use of credit scores in housing and job markets (Rona-Tas 2017). And, as I will explain below in more detail, de-contextualization of data also compromises data quality and the accuracy of profiling. Hence, LoP also aims at ensuring adequate information quality and data processing results.

For sure, no single privacy principle is sufficient to tackle all privacy problems equally. The practical weight of LoP has a lot to do with its exact stipulation.

The more restricted its stipulation is with regard to the limitation to the original purpose, the higher the protection, but the options for future usages are reduced. The more liberal it is, the more flexibility there is, but also greater likelihood of inhering ambiguity with regard to the interpretation of criteria for derogation. Theoretically, there are three variants. Its most restricted form stipulates that data can be used exclusively for the original purpose (variant a). The most liberal form explicitly excludes specific purposes and contexts (variant b), while the more moderate version formulates exceptions from the limitation (variant c). The GDPR, like most regulation, follows variant c.

Limitation of purpose within the GDPR

The GDPR is a comprehensive supranational response to the challenges of balancing power asymmetries in digital information flow. The significance and presence of privacy rules within EU legislation is regarded as high compared to other countries.[4] The regulation does not intend preventing the circulation of data, but aims to achieve that the flow of data does not infringe upon the human right of privacy and data protection (Nicolaidou and Georgiades 2017). The Recital (GDPR, Recital 1) sets out the right to protection of personal data as a fundamental right. Furthermore, it puts an ethical orientation upfront: "The processing of personal data should be designed to serve mankind." (GDPR, Recital 4). The set of privacy principles is stated in Article 5 starting with the claim for lawful, fair and transparent data processing (GDPR, Article 5 (1a)). The limitation on purpose principle comes second. It reads as follows: "[Personal data shall be] collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes […]" (GDPR Article 5 (1b)). There is also a strong correspondence between the EU Directive from 1995 and the five other principles that follow – data minimization, accuracy, storage limitation, integrity and confidentiality, and finally, the accountability of the data controller. Nevertheless, privacy principles were under discussion during the negotiations. A leaked version of an earlier draft of the GDPR proposed by the European Council in which the vigor of LoP was undermined by providing loopholes for incompatible purposes exposed the uncertain status of the principle (Grafenstein 2018, 31). And still, the stipulation of "legitimate" purposes and ruling out "incompatible" data processing is an opening for ambiguity in practice because it can be interpreted differently (see also Bygrave 2014).

Article 6 specifies the operation of the principles. With regard to LoP, it is also a source of further practical challenges. Parts of Article 6 have been criticized for their lack of coherence and lack of an objective scale to determine whether the requirements for circumvention of original purpose are fulfilled, and therefore for the absence of legal certainty (Grafenstein 2018). Article 6 defines the terms for a lawful data processing (and possibly repurposing of data) such as consent given by the data subject, compliance with the legal obligations of the controller, protection of vital interests of the data subject, public interest, and legitimate interests of the controller or third parties (as long as they don't override fundamental rights of the data subject). Furthermore, where processing of data is not based on the data subject's consent it is assessed as being compatible with the initial purpose so long as the interest pursued with the change of purpose outweighs the risks caused by it. Here, the GDPR allows member states to introduce specific provisions for some of those terms to adopt the application of the rule.

In general, the very fact that the member states came to an agreement is regarded as a strong signal that Europe is seeking a balance of responsibility between civil society, market and state (Dijck, Poell, and Waal 2018). Nevertheless, a year after the enactment of the GDPR, evaluations differ substantially between different groups. Enterprise lobbyists point to economic barriers. Data protection advocates indicate loopholes. For example, legal uncertainty in electronic tracking and profiling and in telecommunication services provided across IP networks (over-the-top telecommunication), predominantly the internet, is reported (Schaar and Dix 2019).[5] And EU authorities lament the slowness of corporate compliance, the fragility of enforcement of the rules and the variation in the implementation by the member states.[6] Stricter rules on what constitutes freely given informed consent and the active enforcement of transparency over the extent of data collection are called for in particular.

However, the overall aim of protecting EU citizens from privacy breaches is generally accepted. Moreover, by mobilizing its regulatory capacity the EU shapes policy choices and makes other countries adjust to privacy rules so as to participate in its market. Beyond sanctions and incentives, the European stance on privacy is becoming, as Giovanni Buttarelli, the European Data Protection Supervisor put it, the gold standard and raising the level of privacy protection on a global scale. For instance, the California state government passed the Consumer Privacy Act (CCPA) at the end of 2018, copying many aspects of the GDPR; several other states are working to introduce privacy laws, and calls on US senators to adopt these on the federal level have become louder.[7]

The German debate – Limitation on purpose as safeguard to privacy

The limitation on purpose principle has a noteworthy history for social scientists in the privacy field. The first recorded mention can be found in an expert report of the New York Law Commission in 1965 that identifies fully informed consent as necessary for the revelation of private information but simultaneously characterizes consent as always limited to context and purpose (Ruebhausen and Brim 1965; see also Pohle 2015). These ideas were carried forward in a very influential period in the German data protection debate during the 1970s. The marking of a nexus between privacy and context and the embedding of consent in a purpose frame were at the core of that debate.[8]

Recent literature illustrates the complexity of the debate, resulting in different constructions of privacy (Pohle 2018). Far from following a single line of argument, the German debate was characterized by a lasting struggle over the accurate definition of the social good to be protected and over the related reasoning over phenomena and practices threatening privacy. Briefly summarized, the influential juridical argumentation circled around the question of which right or principle the right to privacy might be derived from.[9] The interpretation of the protected social good moved from the idea of an individual private sphere to privacy as a property of shared social spaces (Podlech 1989). Case-specific policy debates in the 1970s had a formative influence on leading privacy principles and its legal implementation. Three core positions entered data protection legislation during this period. Dealing with the design of a microcensus survey questionnaire, a judicial decision from 1969 stated that it was incompatible with human dignity to completely register and catalogue a person (BVerfG 1969, 6). Secondly, in an advisory report for the German Home Office published in 1972 legal experts amplified the factual scope of a misappropriation rule that had been formulated by the constitutional court before: Personal data should exclusively be processed for those purposes for which it had been collected (Steinmüller 1971). A detail I will come back to in the discussion is that the advisory report regarded LoP as independent from informational consent. And thirdly, in response to debates in preparation of a population census, a judicial decision from 1983 ("Volkszählungsurteil") legally implemented the right to informational self-determination. The decision declared the limitation of data use to its organizational context as a protective goal.

The leading discussion within this formative period had a socio-theoretical nature. For Seidel (1970), who coined the idea of a right to informational self-determination, the right to privacy is manifested in the protection of social engagements and bonds represented in data spaces. Although legal scholars such as Seidel dominated the debate, there was a short period of interdisciplinary exchange on privacy between 1972 and 1978 (Pohle 2018). Sociologists participating in the debate took up context as a key concept and applied it to a definition of privacy, with reference to prominent theoretical schools such as symbolic interactionism, role theory and social system theory. The claims of context-orientated sociology that social actions and expressions pointed to situations were adopted by role theory in which the exchange of information was regarded as context specific behavior. According to this theory, different information interests are regarded as tied to different functional roles. From this perspective, information exchange related expectations contribute to the stabilization of role structures and the social system as a whole (Parsons 1951). Drawing on role theory, Müller and Kuhlmann defined privacy as "the individual's 'visibility' in varying contexts" (Müller und Kuhlmann 1972, 590). By that, they went beyond the common distinction between a private and a public sphere on which earlier privacy concepts were based. Pointing to the "the role-specific exclusivity of information" they allowed for privacy entitlements in public contexts (ibid, 595; see also Pohle 2018). Another twist was to use purpose and context as a looking glass to determine the sensitivity of data rather than the content of data (Lenk 1973; see also Miller 1969).

The boundedness to context was implemented as limitation on purpose within the German legislation. In an influential summary and interpretation of the debate Hoffmann (1991) argued that LoP is a prerequisite for informational self-determination when participating in public social life. He particularly stressed the threats of misappropriation of information through automated data processing. The subject matter of protection is no longer a type of data but context and the purpose of use targeted by the data subject. In particular, privacy was understood as valuable not just to the individual but to the community as a whole (Podlech 1989). Privacy is a quality of the way the communal information exchange is organized. Therefore, a toleration of de-contextualization of information harms communal exchange in general. For Hoffmann LoP is an equivalent to the preservation of context with regard to the targeted use (Hoffmann 1991). Therefore, LoP is regarded as the key mechanism to guarantee privacy.

These arguments from decades ago still pin down the core problem of privacy: the appropriate distribution of information. The socio-theoretical turn to the recognition of privacy as linked to participation

in public social life advanced a position, which again appears topical for the current challenges of information technologies. In this simplified historic reading, the accomplishment of purpose and context limitation is that they provide a criterion to keep different information (or communication) domains separate. The linkage of this theorizing on data protection to the theory of functional differentiation opens up a perspective beyond domains of information (Rost 2013). It allows seeing privacy as construct of modern society, an invention to justify the functional differentiation of information. An information industry, which has an increasing potential to intrude in context embedded activities and integrate and cross-reference data files that are deprived of context limitation, is making everything visible. This can be viewed as a somewhat newly generated "village situation" in which everyone knows everything about everybody else. However, it is not simply a regress to a pre-modern segmented social order because of a unique asymmetry. The intermediating institutions themselves are beyond scrutiny. Compared to a platform such as Facebook, in a village people meet at public spaces (in a modern village this would be the church, the pub or the market). The priest or the shaman might know a bit more than others about the villagers simply because of their roles. Compared to Google, in a village the stories of the villagers are recorded by the elder (the modern village might have a library). But all these positions are under public scrutiny and can be held accountable for what they do with their knowledge. The big digital platforms use their data without the public being privy to it. Privacy regulation balances this feature of the technological infrastructure. LoP is key for this maintenance of functional differentiation because it explicitly signifies the role character of information. LoP safeguards privacy and makes "the village" a city.

How de-contextualization jeopardizes accuracy

Through de-contextualization contexts disappear in different ways. There are two types. In a first variant, data is moved from one realm to another. What gets lost here is the meaning of the data created in the original context and shaped by its intended use. A methodological critique of this problem was articulated in the German privacy debate. It was reasoned that misappropriation of data carried the risk of distorted meaning. Different arguments were brought into the debate. One was context-related ambiguity of meaning. Literature drawing on symbolic interactionism adjusted the focus from information to communication and argued that communication is not fully comprehensible when set outside its context, situation or social relation (Rüpke 1976). From this perspective privacy was to be understood as a shield against misunderstanding and false interpretation. The literature investigating administrative mass data identified the bracketing of context of data origin as a main source for error (Bick and Müller 1983).

A second form of de-contextualization concerns measurement and quantification. Calculative practices must drop information to make cases comparable and to fit them into categories. Here de-contextualizing means ignoring unique or relational characteristics. At the same time, the categories become essentialized. It is overlooked that classifications are dependent on the blurring of heterogeneity and on the enforcing of differences (Boltanski and Thévenot 1983, Zeruvabel 1991, 1996, Bowker and Star 2000), and that they make invisible the interventional character of measurement they depend on (Thévenot 1984, 2009, Porter 1995, Diaz-Bone and Didier 2016).

In the digital world both de-contextualizations tend to co-occur. Quantifying and categorizing over different data sources from very different contexts is the case for quite many digital data usages. Both variants of de-contextualization impact the accuracy of information to different degrees and affect the appropriate use of data unless they become re-contextualized.

Although the debate on data accuracy and context is apparently not novel, claims emerging with the proliferation of information technologies and big data methods make it highly topical (Lewis 2015, Marres 2017). These technologies lead to a new idea of "traceability" of social life, which often identifies data as facts. An often-cited assumption of contemporary data practices is that "with enough data, the numbers speak for themselves" (Anderson 2008). The faith in data can be observed in the commercial field and even in academia. It is the vision of computational social science that compiled data adequately explains the world and helps to achieve a comprehensive picture of patterns of individual and group behavior (Lazer et al. 2009). The main objections against the "data as fact" claim are with reference to context (Edwards et al. 2011). Collection and extraction of data never covers all information available. Usually they are themselves embedded in an institutional context and follow a specific purpose that determines the choices and decision throughout the process. This is nothing specific to digital data but is a general property of mass data (Baur 2009). Choices and interpretations through data collection are most often purpose-driven. This also means "different people in different contexts with different

goals will choose different answers as they construct their data models" (Shaw 2015, 3). At the same time data is continuously repurposed (Andrejevic and Gates 2014).

The powerful effect of complex mass data comes from the aggregation of different data sources. However, the literature increasingly points to the challenges (and traps) in the way mass-produced digital data is processed. The key problem is veracity, namely, that "data are not generated from instruments and methods designed to produce valid and reliable data amenable to scientific analysis" (Japec et al. 2015, 849). Data used and transformed into data sets starting with the original source and ending in data warehouses are often by-products of other processes. Here we observe mostly de-contextualization of type one. When datasets are merged a series of processes take place. Data is reduced, parts of data are extracted and transformed into new variables by cleaning, aggregating, reformatting, recoding, matching records. These transformative steps rely heavily on technically complex processing (data mining, algorithms) and involve a high level of data interpretation (Japec et al. 2015). Due to the underlying assumptions about data along these steps, which are often not systematically reflected, literature talks about transformation biases (Baker 2017). Transformations don't take ambiguity of meaning into account, question data validity and jeopardize accuracy. Other concerns question if these data actually measure natural behavior and point to the artificiality of platform designs. The specific configuration of software interfaces suggests certain actions and limits choices (Shaw 2015). Again other methodological concerns touch on the representativity of data. There is a systematic selection bias because some parts of the population are simply not online (population bias). Also, there are most certainly "holes" in individual data records. The handling of missing data in complex databases either by imputation or fusion techniques also runs the risk of reducing accuracy. In survey designs these common sources for error are systematically controlled for. For big data analysis they pose even bigger challenges (Baker 2017).

This leads us to de-contextualization of type two. What is relevant to data, is also relevant to the usage of statistical profiles. Those are based on data driven classifications on the assumption that digital infrastructures depict invisible patterns in society and "that we can know what people are doing in an objective manner, without biases, without lying, without kidding ourselves, of trying to present a different image than what we are" (Barabási 2012). However data science has to be aware of the (natural) boundary and measurement fallacies (Krenn 2017) discussed above. The objective appearance of classifications gives them a strong legitimizing push for its usage. This insight is particularly relevant for complex mass data that also carries algorithmic bias (Crawford 2013). All these threaten the validity of data.

From this follows that complex mass data only produce valid results for appropriate contexts and require complex interpretation. The collection, as well as the aggregation of different data sources, demand special care to preserve data context. What kind of knowledge may be gained from digital mass data is a question that has to be discussed elsewhere. However, no matter what kind of data driven real world decision is made or how scientific data is used, safeguarding context preserves the pragmatic meaning that individuals attach to their own behavior. In other words, LoP is equally a protection against misinterpretation and distortion of the pragmatic meaning of participation in the digital community.

Discussion

The distribution of data remains the present and future challenge of privacy. Information technologies and supporting infrastructures build the substrate for tracking, compiling and classifying data. The design of these technologies and applications is highly asymmetrical regarding the way the exchange of information is organized and becomes comprehensible. Data protection regulation attempts to balance this asymmetry and to protect the weaker party, the individual user, who is exposed to these technologies unless they abstain from participation in digital services. Looking back to early discussion showed that from the early days of the development of information technology LoP and context attachment have been considered as principles to safeguard such values as privacy and accuracy. They provided an answer to the question of how to assess the appropriateness of data access and distribution. Since then de-contextualization and disrespect of targeted purposes mark the misappropriation of data. De-contextualization and disrespect of purpose define a violation of privacy and as a harm to accuracy they present a distortion of information. Hence, informational norms grounded in context are not just only a 21st century invention (Nissenbaum 2009), they might just still provide answers to contemporary challenges of informational asymmetry and be a valid guide for identifying privacy violations and false interpretations.

Pondering the implications of LoP brings us back to the various forms it can take. It is clear that the more liberal its stipulation is towards derogations from the intended purpose, the lower is the de facto level of protection. I would like to discuss this looking

at individual consent as basis of legitimacy for repurposing of data. From a theoretical viewpoint LoP is not necessarily intertwined with individual consent, as the narrative of the German debate has shown. In practice, consent often results in a potential loss of context. Let us picture this. In order to comply with GDPR requirements internet service providers have to obtain consent for data processing. In everyday practice this means that websites or apps often prompt data subjects to consent to quite hazy future data processing. For instance, food delivery platforms ask users for their consent to cookies that identify which restaurants they like, what food they prefer and where and when they like to have their meals. Moreover these platforms prompt exchange of data with third-party suppliers such as social media sites to personalize information. The recent retreat of the food delivery service Deliveroo from Germany exposed a serious question: What actually happens to such data, obtained with consent for such imprecise purposes, when the company goes bankrupt? Who hinders the liquidators from selling it for completely different uses? Also, specific configurations of app permissions are an opening for service providers to work around LoP. The majority of users consent to share digital trace data such as geolocations, app usage and access to contact lists. As a recent app-study showed, users hardly differentiate between the different data requests (Kreuter et al. 2018).

Another evocative example gives a recent report by Privacy International, which reveals that mental health websites in France, Germany and the UK shared information on depression with third parties (Privacy International 2019). This included information on web searches and depression test results. This is a serious privacy violation considering the impact it might have on profiling. In addition to undesired personalized advertising, such data could seriously affect major future decision processes in the job market or in other domains. For this reason, health data already belong to a special category within the GDPR and merit higher protection (GDPR, Recital 53). However, this targets the national health sector, and mental health websites are privately operated platforms. Most websites contained third-party elements such as tracking cookies or java script, making devices identifiable and saving data on website activities. Many of the observed websites didn't meet the GDPR standards for freely given, specific, informed and unambiguous consent with a clear affirmative action (such as a GDPR conform cookie banner). So, this might appear counterintuitive as an example of the consent requirement. However, as the study mentioned above showed, users hardly differentiate between consent requests. Hence, compliance might only be a part of the problem. This case also raises serious concerns about the qualification of consent the way it is implemented on most website as safeguard against overriding fundamental rights of individuals. Without doubt, consent is an important feature for information exchange. But it is debatable if consent alone should always be sufficient for deliberately repurposing data. This story about mental health websites demonstrates how important the specificity and context of information exchange are as basic principles. In the case of mental health websites an exception from LoP does generally not appear appropriate.

Another strategy of websites is to link consent requests with functionality incentives. Consent becomes a condition of using the website's services. These are just a few examples for modes of industrial data processing that use (more or less) informed consent to repurpose data in everyday practice. Not all purposes for which data usage is consented correspond with contextual meaning of digital traces and purpose. Of course, it is not always easy to determine what the purpose of the data is. And it is even harder to define once and for all what a good or bad use of data is. Given that almost all mass data from platforms or applications have to deal with this tension between the intended visibility of the user in an exclusive context and the translation of data to other purposes, be it consensual or not, the discussion on implementation of privacy rules will continue.

Considering the potential social impact profiling has on users, a more restricted form of LoP appears better qualified for balancing the power asymmetry between organizations and the individual user. It is worthwhile thinking about earmarking exclusive purposes for data processing as a feasible option – at least for some information domains. Implemented in such a way, LoP could become an even stronger anchor for testing and preserving the controllability of data flow. It might also allow dealing with situations where users are not aware of providing data and their consent is not asked for.

Conclusion

Many productive ideas fall into oblivion only to later experience a renaissance. Context appears to be just such an old concept that still provides answers to contemporary questions. The problem of participation and privacy in the new public informational realm is a contemporary challenge for the ordering of democratic societies. The strength of a context perspective is that it covers the distribution as well as the accuracy of data. Hence, limitation on purpose as the prime privacy principle has the potential to cover the core matters

for the regulation of information infrastructures. The limits of the imaginaries behind LoP as principle to control processes of information exchange are in its legal (and technical) implementation. National varieties show that things can be different. Any concrete construction of privacy has to prove its potential to live up to transnational demands. The GDPR provides a legal basis for Europe on the key principle of LoP. Still, its impact is limited to the conclusiveness of bringing context back into the everyday use of data.

Endnotes

1 Regulation (EU) 2016/679 of the European Parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

2 Sanctioning mechanisms of regulatory policies have also been strengthened. Breaches of the GDPR can be fined up to 4% of a firm or organizations' annual global turnover, which generates a strong incentive for compliance. An example is the record-setting fine imposed on Google in January 2019 (not to be confused with the anti-trust fine in July 2018) by French data protection authorities for illegal practices on mobile devices.

3 For example, only a minority of smartphone apps correctly declare data sharing policies. Privacy breaches are particularly serious in, for example, health apps passing on information on depression or smoking habits to Facebook or Google (Huckvale et al. 2019).

4 There are varying explanations given. From a socio-economic perspective the formation of data privacy regulation was influenced by the interplay of domestic policies regarding the consumer lending sector and transnational post-war globalization policy activism (Trumbull 2011). Institutionalist arguments focus on the leading role of national privacy authorities and regulatory institutions (Newman 2008); from the 1970s on they promoted privacy concerns at the European level through networks and coercive power. And another strand of literature follows more a cultural argument, seeing privacy standards as a reflection of deep-seated national values (Bellman et al. 2004). A recent continuation of the latter is given by an assessment of European policies (GDPR) as a sign for the upholding of "public values in a connective world" (Dijck, Poell, and Waal 2018).

5 The regulation of user tracking demonstrates the difficulties regarding the national implementation of the GDPR. Looking at Germany, there are different interpretations on the question of which guidelines to administer. Data protection agencies interpret the GDPR as overruling national law, which allows user tracking (Schaar and Dix 2019).

6 The national implementation of the GDPR gives countries enough scope to be an obstacle to the intention of the regulation as criticized by Verá Jourová, the European Commissioner for Justice. Speech on the occasion of the first anniversary of the GDPR. http://europa.eu/rapid/press-release_SPEECH-19-2697_en.htm (Last access September 7th 2019)

7 https://www.cnbc.com/2019/05/23/gdpr-one-year-on-ceos-politicians-push-for-us-federal-privacy-law.html (Last access September 7th 2019) https://www.nbcnews.com/tech/tech-news/california-bringing-law-order-big-data-it-could-change-internet-n1005061 (Last access September 7th 2019)

8 The German debate was in turn influenced by the debate in the US that took a lead role in the privacy debate (Pohle 2018). Likewise it's not only the German discourse that regards context. For instance, Brenton (1964) had already called attention to the risk of de-contextualization of private information through computer technology.

9 An early German source mentioned is Kohler (1880), who described the right to privacy as a fundamental individual right by the end of the 19th century. Shortly after, a first reference to privacy was published in the US by Warren and Brandeis (1890).

References

Acemoglu, Daron. 2002. "Technical Change, Inequality, and the Labor Market." *Journal of Economic Literature* 40 (1): 7–72.

Anderson, Chris. 2008. "The end of theory: the data deluge makes the scientific method obsolete". In *Wired*. Source: https://www.wired.com/2008/06/pb-theory/ (last access June 23rd 2018)

Andrejevic, Mark, and Kelly Gates. 2014. "Big Data Surveillance: Introduction." *Surveillance and Society* 12 (2): 185–196.

Baker, Ray. 2017. "Big Data. A Survey Research Perspective." In *Total survey error in practice*, edited by Paul P. Biemer, Edith de Leeuw, Stephanie Eckman, Brad Edwards, Frauke Kreuter, Lars E. Lyberg, N. Clyde Tucker und Brady T. West, 47–69. New York: John Wiley & Sons, Inc.

Barabási, Laszlo. 2012. "Thinking in Network Terms." In *Edge*. Online Source: https://www.edge.org/conversation/albert_l_szl_barab_si-thinking-in-network-terms (Last access May 23rd 2019)

Baur, Nina. 2009. "Measurement and Selection Bias in Longitudinal Data. A Framework for Re-Opening the Discussion on Data Quality and Generalizability of Social Bookkeeping Data." *Historical Social Research* 34 (3): 9–50.

Bellman, Steven, Eric J. Johnsons, Stephen J. Kobrin, and Gerald L. Lohse. 2004. "International Differences in Information Privacy Concerns: A Global Survey of Consumers." *Information Society* 20: 313–324.

Bick, Wolfgang, and Paul J. Müller. 1984. „Sozialwissenschaftliche Datenkunde für prozeßproduzierte Daten: Entstehungsbe-

dingungen und Indikatorenqualität." In *Sozialforschung und Verwaltungsdaten*, edited by Wolfgang Bick, Reinhard Mann, and Paul J. Müller, 123–159. Stuttgart: Klett-Cotta.

Bowker, Geoffrey C., and Susan L. Star. 2000. *Sorting Things Out. Classification and Its Consequences*. Cambridge, MA: The MIT Press.

Brenton, Myron. 1964. *The Privacy Invaders*. New York: Coward-McCann Inc.

Brynjolfsson, Erik, Andrew McAfee, and Michael Spence. 2014. "New World Order: Labor, Capital, and Ideas in the Power Law Economy." *Foreign Affairs* 93 (4): 44–53.

Bygrave, Lee A. 2014. *Data Privacy Law. An International Perspective.* New York: Oxford University Press.

Christl, Wolfie, and Sarah Spiekermann. 2016. *Networks of Control. A Report on Corporate Surveillance, Digital Tracking, Big Data and Privacy.* Wien: Facultas.

Datta, Amit, Michael C. Taschantz, and Anupam Datta. 2015. "Automated Experiments on Ad Privacy Settings. A Tale of Opacity, Choice and Discrimination." *Proceedings on Privacy Enhancing Technologies* 2015 (1): 92–112. DOI: https://doi.org/10.1515/popets-2015-0007

Diaz-Bone, Rainer, und Emmanuel Didier (Ed.). 2016. *Conventions and Quantification – Transdisciplinary Perspectives on Statistics and Classifications* (Special Issue). *Historical Social Research* 41 (2).

Dijck, José van, Thomas Poell, and Martijn de Waal. 2018. *The Platform Society: Public Values in a Connective World.* New York: Oxford University Press.

Edwards, Paul, Matthew S. Mayernik, Archer Batcheller, Geoffrey Bowker, and Christine L. Borgman. 2011. „Science friction: Data, metadata, and collaboration." *Social Studies of Science* 41 (5): 667–690.

Fourcade, Marion, and Kieran Healy. 2013. "Classification situations: Life-chances in the neoliberal era." *Accounting Organizations and Society* 38 (8): 559–572.

Gandy, Oscar H. Jr. 1993. *The panoptic sort: A political economy of personal information. Critical studies in communication and in the cultural industries*. Boulder, CO. Westview Press.

Grafenstein, Maximilian von. 2018. *The Principle of Purpose Limitation in Data Protection Laws. The Risk-based Approach, Principles, and Private Standards as Elements of Regulating Innovation*. Baden-Baden: Nomos Verlagsgesellschaft.

Hoven, Jeroen van den. 2001. „Privacy and the Varieties of Informational Wrongdoing." In *Readings in Cyberethics*, edited by Richard A. Spinello and Herman T. Tavani, 488–500. Sudbury, MA: Jones and Bartlett Publishers.

Huckvale, Kit, John Torous, and Mark E. Larsen. 2019. "Assessment of the Data Sharing and Privacy Practices of Smartphone Apps for Depression and Smoking Cessation." *JAMA Network Open* 2 (4): e192542. Onlinesource: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2730782
(Last access September 7th 2019)

Kohler, Josef. 1880. *Das Autorrecht*. Jena: Verlag Gustav Fischer.

Krenn, Karoline. 2017. "Markets and Classifications – Constructing Market Orders in the Digital Age. An Introduction." *Historical Social Research* 42 (1): 7–22.

Kreuter, Frauke, Georg-Christoph Haas, Florian Keusch, Sebastian Bähr, and Mark Trappmann. 2018. "Collecting Survey and Smartphone Sensor Data With an App: Opportunities and Challenges Around Privacy and Informed Consent." *Social Science Computer Review*. DOI: https://doi.org/10.1177/0894439318816389 (Last access September 7th 2019).

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert Laszlo Barabasi, Devon Brewer, et al. 2009. "Life in the network: the coming age of computational social science." *Science* 323 (5915): 721–723.

Lenk, Klaus. 1973. „Datenschutz in der öffentlichen Verwaltung." In *Datenschutz. Beiträge zur juristischen Informatik Band 1*, edited by Wolfgang Kilian, Klaus Lenk, and Wilhelm Steinmüller, 15–50. Frankfurt a.M.: Athenäum-Verlag.

Lewis, Kevin. 2015. "Three fallacies of digital footprints." *Big Data and Society* 2 (2): 1–4.

Liang, Fan, Vishnupriya Das, Nadia Kostyuk, and Muzammil M. Hussain. 2017. "Constructing a Data-Driven Society as a State Surveillance Infrastructure." *Policy and Internet* 10 (4): 415–453.

Japec, Lilli, Frauke Kreuter, Marcus Berg, Paul Biemer, Paul Decker, Cliff Lampe, Julia Lane, Cathy O'Neil, Abe Usher. 2015. „Big data in survey re-search. AAPOR task force report." *Public Opinion Quarterly* 79 (4): 839–880.

Makri, Eleni-Laskarini and Costas Lambrinoudakis. 2015. "Privacy Principles: Towards a Common Privacy Audit Methodology." *Lecture Notes in Computer Science* 9264: 219–234.

Marres, Noortje. 2017. *Digital Sociology. The Reinvention of Social Research*. Cambridge and Malden: Polity Press.

Mendoza, Isak, and Lee A Bygrave. 2017. "The Right Not to Be Subject to Automated Decisions Based on Profiling". In *EU Internet Law: Regulation and Enforcement,* edited by Tatiani Synodinou, Philippe Jougleux, Christiana Markou, and Thalia Prastitou, 77–98. Springer International Publishing AG. Doi: https://doi.org/10.1007/978-3-319-64955-9_4

Miller, Arthur R. 1969. "Personal Privacy in the Computer Age: The Challenges of a New Technology in an Information-Oriented Society." *Michigan Law Review* 67 (6): 1089–1246.

Müller, Paul J., and H.H. Kuhlmann. 1972. "Integrated information bank systems, social book-keeping and privacy." *International Social Science Journal* 24 (3): 584–602.

Newman, Abraham l. 2008. *Protectors of Privacy. Regulating Personal Data in the Global Economy.* Ithaca and London: Cornell University Press.

Nicolaidou, Irene L., and Constantinos Georgiades. 2017. "The GDPR: New Horizons." In *EU Internet Law. Regulation and Enforcement,* edited by Tatiani Synodinou, Philippe Jougleux, Christiana Markou, and Thalia Prastitou, 77–98. Springer International Publishing AG. Doi: https://doi.org/10.1007/978-3-319-64955-9_1

Nissenbaum, Helen. 2015. "Respect for context as a benchmark for privacy online: What it is and isn't." In *Social Dimensions of Privacy. Interdisciplinary Perspectives*, edited by Beate Roessler, and Dorota Mokrosinska, 278–302. Cambridge: Cambridge University Press.

Nissenbaum, Helen. 2009. *Privacy in Context. Technology, Policy, and the Integrity of Social Life.* Stanford: Stanford University Press.

OECD. 2011. Thirty Years After. The OECD Privacy Principles. Onlinesource: http://www.oecd.org/sti/ieconomy/49710223.pdf (Last access August 29th 2019)

Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Parsons, Talcott. 1951. *The Social System.* New York: Free Press.

Podlech, Albert. 1989. „Die Grundrechte. Art. 2 Abs. 1." In *Kommentar zum Grundgesetz für die Bundesrepublik Deutschland Band 1*, edited by Erhard Denninger, Helmut Ridder, Helmut Simon, and Ekkehard Stein, 266. Neuwied: Luchterhand Verlag.

Pohle, Jörg. 2018. Datenschutz und Technikgestaltung: Geschichte und Theorie des Datenschutzes aus informatischer Sicht und Folgerungen für die Technikgestaltung. Berlin: Humboldt-Universität zu Berlin. Onlinesource: https://edoc.hu-berlin.de/handle/18452/19886 (Last access September7th 2019)

Pohle, Jörg. 2015. „Zweckbindung revisited." *DANA – Datenschutz Nachrichten* 38 (3): 141–145.

Poon, Martha. 2016. "Corporate Capitalism and the Growing Power of Big Data: Review Essay." *Science, Technology, and Human Values* 41 (6): 1088–1108.

Porter, Theodore M. 1995. *Trust in numbers : the pursuit of objectivity in science and public life.* Princeton, N.J.: Princeton University Press.

Privacy International. 2019. *Your mental health for sale. How websites about depression share data with advertisers and leak depression test results*. Onlinesource: https://privacyinternational.org/sites/default/files/2019-09/Your%20mental%20health%20for%20sale%20-%20Privacy%20International.pdf (Last access September 7th 2019).

Rost, Martin. 2013. "Zur Soziologie des Datenschutzes." *DuD – Datenschutz und Sicherheit* 2: 85–91.

Rona-Tas, Akos. 2017. "The Off-Label Use of Consumer Credit Ratings." *Historical*
Social Research 42 (1): 52–76. Doi: https://doi.org/10.12759/hsr.42.2017.1.52-76

Ruebhausen, Oscar M., and Orville G. Brim. 1965. "Privacy and Behavioral Research." *Columbia Law Review* 65 (7): 1184–1211.

Rüpke, Giselher. 1976. *Der verfassungsrechtliche Schutz der Privatheit. Zugleich ein Versuch pragmatischen Grundrechtsverständnisses*. Baden-Baden: Nomos Verlagsgesellschaft.

Schaar, Peter, and Alexander Dix. 2019. *Datenschutz im digitalen Zeitalter: Umsetzung der Datenschutzgrundverordnung (DSGVO) – Bilanz ein Jahr nach Inkrafttreten.* Gutachten im Auftrag der Fraktion Bündnis 90/Die Grünen im Bundestag. Onlinesource: https://www.gruene-bundestag.de/fileadmin/media/gruene-bundestag_de/themen_az/datenschutz/PDF/Gutachten_DS-GVO.pdf (Last access September 7th 2019)

Seidel, Ulrich. 1970. "Persönlichkeitsrechtliche Probleme der elektronischen Speicherung privater Daten." *Neue Juristische Wochenschrift* 23: 1581–1583.

Sevignani, Sebastian. 2017. "Surveillance, Classification, and Social Inequaliy in Informational Capitalism: The Relevance of Exploitation in the Context of Markets in Information." *Historical Social Research* 42 (1): 77–102. Doi: https://doi.org/10.12759/hsr.42.2017.1.77-102

Shaw, Ryan. 2015. "Big data and reality." *Big Data and Society* 2 (2): 1–4.

Steinmüller, Wilhelm, Bernd Lutterbeck, Christoph Mallmann, U. Harbort, G. Kolb, and J. Schneider, Peter. 1971. *Grundfragen des Datenschutzes. Gutachten im Auftrag des Bundesministeriums des Inneren*, BT-Drucksache VI/3826, Anlage 1.

Thévenot, Laurent. 1984. "Rules and implements: investment in forms." *Social Science Information* 23 (2): 1–45.

Trumbull, Gunnar. 2011. "Between Global and Local: The Invention of Data Privacy in the United States and France." In *The Voice of the Citizen Consumer. A History of Market Research, Consumer Movements, and the Political Public Sphere*, edited by Kerstin Brückweh, 199–224. New York: Oxford University Press.

Warner, Malcolm, and Michael Stone. 1970. *The Data Bank Society: Organizations, Computers and Social Freedom*. London: George Allen and Unwin Ltd.

Warren, Samuel D., and Louis D. Brandeis. 1890. "The Right to Privacy." *Harvard Law Review* 4 (5): 193–220.

Zerubavel, Eviatar. 1991. *The fine line. Making distinctions in everyday life.* Chicago: The University of Chicago Press.

Zerubavel, Eviatar**.** 1996. "Lumping and Splitting: Notes on Social Classification." *Sociological Forum* 11 (3): 421–433.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Front*. New York: Public Affairs.