Gunnar Stevens / Paul Bossauer*

Dealing with Personal Data in the Age of Big Data Economies

The technological development of the digital computer and new options to collect, store and transfer mass data have changed the world in the last 40 years. Moreover, due to the ongoing progress of computer power, the establishment of the Internet as critical infrastructure and the options of ubiquitous sensor systems will have a dramatic impact on economies and societies in the future. We give a brief overview about the technological basics especially with regard to the exponential growth of big data and current turn towards sensor-based data collection. From this stance, we reconsider the various dimensions of personal data and and market mechanisms that have an impact of data usage and protection.

I. Technical basics

The digital computer represents the technical foundations of today's data economy with its capabilities to collect, process, store, and transfer data easily. The digital computer is not new, but has nearly 100 years of history. Moreover, the general construction has not changed significantly in recent decades. Still, the computer power has dramatically improved in that time.

Gordon Moore, co-founder of INTEL, was one of the first who recognized these dynamics. In 1965, he observed that the performance of computers is doubling every two years. Within the same period of time, the costs for the same capacity are halved.[1] Moore's law – or more precisely Moore's prediction – was valid the last five decades and it seems that it will continue for the next few years. Moreover, the *law* essentially holds for all performance indicators mentioned in Table 1 such as processing power, memory size and speed and data transfer rate.

* Prof. Dr. Gunnar Stevens is Divisional Director for IT-security and consumer computing at the University of Siegen and owner of the chair for business administration and business informatics at the University of Applied Sciences Bonn-Rhein-Sieg. Paul Bossauer is research associate at the University of Applied Sciences Bonn-Rhein-Sieg.

1 Roberts, "Beyond Moore's law: Internet growth trends." Computer 33.1 (2000): 118.

ZGE / IPJ 9 (2017), 266–278 DOI: 10.1628/186723717X15069451170856
ISSN 1867-237X

Table 1 Basic building blocks of IT systems and infrastructures

Component	*Description*	*Performance Indicators*
Data Processing – CPU	Executing logical and arithmetic operations	– (Floating Point) Operations per Second [FLOPS][2]
Data Storage – RAM – HD / SSD	Storing data temporarily (RAM) or permanently (HD / SSD)	– Memory Size – Access Time – Data transfer rate
Data Transfer – Bus – Networks	Transferring data within and among computers	– Access Time – Data transfer rate
Data Collection – Sensors – User Interfaces	Converting analogue signals of the environment to digital data and vice versa	– Type of sensor – Sampling rate – Size

An impressive example presents the progress made by the storage capacity. In 1956, three men were required to transport a hard drive of just under 4 MB (Figure 1 left) – a ridiculously small capacity, just enough to save a holiday snapshot of a modern digital camera. In contrast, with our Smartphone we transport 10,000 times more memory in our pocket everyday – often without being aware of it (Figure 1 right). This progress was also reflected by the memory cost. For instance, Figure 2 shows the decrease of a gigabyte memory from 1,000,000 $ in the 1980s to 0.10 $ today.

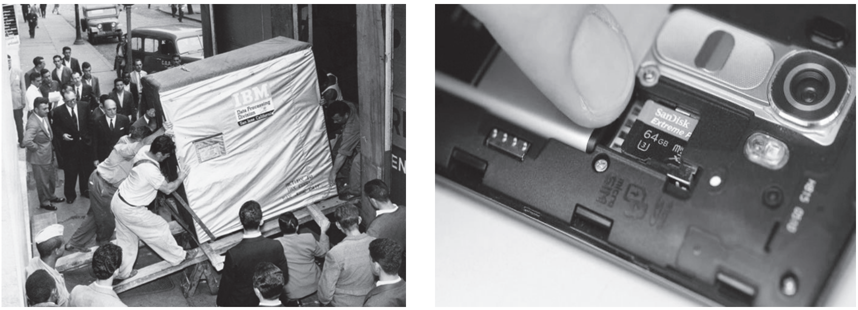

Figure 1: The memory size in the early days of computers and today: The IBM 350 Hard Disk in the year 1956 with 4 MB memory size (left). Today, even micro SD memory cards can store 10,000 times as much as the first hard disk (right).

The same applies to the computing performance as well as to transmission rates of computer networks. This development is further enhanced by the fact that

2 Roughly spoken FLOPS (Floating Point Operations per Second). This indicator is given by multiplying the numbers of CPU cores with the clock frequency of the CPU. See: Wu, Performance evaluation, prediction and visualization of parallel systems. Vol. 4. Springer Science & Business Media (2012): 116.

more and more computers are being produced so that the overall world capacity grows even faster. Hilbert and López, for instance, estimate that humankind was able to store about 300 exabytes (300,000,000,000, 000,000,000 bytes), communicate about 2 zettabytes (2,000,000,000,000,000,000,000 bytes) and carry out about 6 ExaFLOPS (600,000,000,000, 000,000,000 FLOPS) on general-purpose computers in 2007.[3] McAfee and Brynjolfsson use the picture that in the internet every second more data is transferred than were stored in the entire internet just 20 years ago.[4]

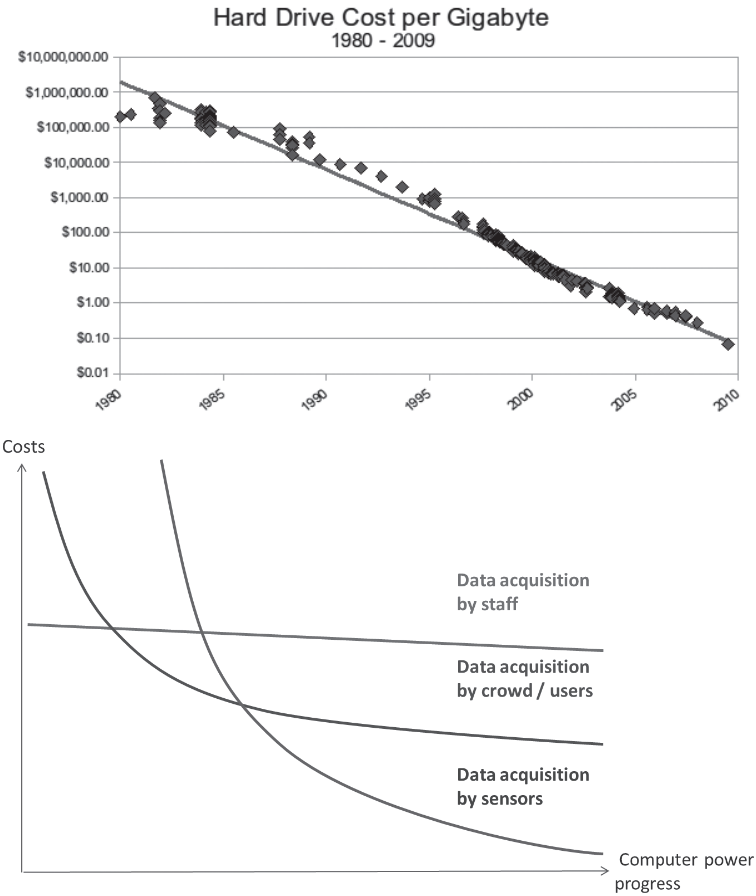

Figure 2: Top: Storage Cost per Gigabyte.[5] Bottom: Cost degression trough IT for data acquisition

3 Hilbert and López, "The World's Technological Capacity to Store, Communicate, and Compute Information," Science 332, no. 6025 (2011): 60.

4 McAfee and Brynjolfsson, "Big Data: The Management Revolution," Harvard Business Review 90, no. 10 (2012): 4.

5 Source: http://www.mkomo.com/cost-per-gigabyte [Access date: 8/2/17].

Currently, there is also a change in the way how data will be collected: Very briefly, three types of data collection can be distinguished: firstly, manually entering data by paid staff; secondly, generating data by the crowd; thirdly collecting data by sensors.

The first type is characterized by traditional ways of data acquisition, e. g. in the process creating an encyclopedia as well as managing a business transaction. Both have in common that staff was paid to generate the needed data – in the case of encyclopedias e. g. high-skilled experts were paid to write articles, in the case of ERP systems e. g. poorly-trained workers were paid to make the inventory or the checkout. Hence, we could label this age as the professional data production.

The second type is characterized by social innovations of self-service processes and user-generated content. The shift was supported by the Internet which made Web 2.0 as well as E-Commerce possible. Here, customers and users were much more responsible to create the needed information. The most impressive example is Wikipedia, where unpaid users write a digital encyclopedia collaboratively. Concerning this, the users' daily use of web shops, fill out forms as well as clickin on articles and writing user reviews seems less impressive – still it comes along with an excessive cost reduction for companies when outsourcing business processes completely or partially to the customers, e. g. by product rating by consumers.[6]

The third type is characterized by digital sensors. The most known example is GPS built into every smartphone. However, there are many of other sensors such as RFID, NFC, infrared, pressure or temperature sensors, tiny cameras and microphones, smart meters, etc. Moreover, digital sensors have greatly improved measurement accuracy and became smaller and cheaper in the last years.

Because of economic reasons digital sensors will be the future of data collection (cf. Figure 2). While in the past data was collected by humans (either by paid or unpaid customers or users) using user interfaces to type in text and numbers, we observe the dissemination of the Internet of Things where an exponential growth of digital sensors becomes responsible for the data acquisition.

In summary, the outlined exponential growths in combination with progresses in data mining, business analytics and machine learning enables and drives the recent trend of *data economies* magnifying the size and augmenting the scope of big data[7] concerning storage volume that enables to run big data centers at low costs, concerning computing performance that enables big data analysis even with more complex pattern recognition and deep learning algorithms and concerning advanced computer networks that enables the transport of locally collected data to central servers in the cloud.

6 Kleemann et al., "Un(der)paid Innovators: The Commercial Utilization of Consumer Work through Crowdsourcing," Science, Technology & Innovation Studies 4, no. 1 (2008): 14.

7 Acquisti, "From the Economics of Privacy to the Economics of Big Data," in Privacy, Big Data, and the Public Good: Frameworks for Engagement, Cambridge University Press (2014): 76.

Last but not least, the network of billions of digital sensors reduces the costs for the data collection dramatically. Because of this, it is getting economically attractive to set up large data streams and data stocks with high quality data. These streams of sensor data are typically merged with user-generated and paid data stocks e. g. to create fine-grained consumer profiles.[8] Using all this data in combination allows a cheap and ubiquitous observation of the surrounding and the actual context.

II. Data Economies

1. Dimensions of Big Data

"The big data movement, like analytics before it, seeks to glean intelligence from data and translate that into business advantage. However, there are three key differences [concerning volume, velocity and variety]."[9]

The new technical possibilities for big data stocks are the key driver for emerging data markets that asks for new business models as well as new infrastructures such as platforms for selling, buying, or trading data. These platforms are also called data markets and supply analysts, business applications, and developers with meaningful data and an eased data access.[10] Such data markets enable access for multiple users to the available data under different regulations and variant licensing strategies.[11] They rely on a fusion of various data sources, which could be internally stored in ERP and CRM systems, public from the web, data provided by government or generated manually from consumers or automatically from sensors.[12]

In general, three types of business models can be distinguished: Firstly, data stocks can be used internally to optimize existing processes and improve the quality of products and services. Within companies, the most common use is to better understand and reach out to consumers, e. g. using data analytics to predict consumers' willingness to pay or using this information for personalized pricing.[13] Secondly, companies can benefit from creating new data-driven services and business models. A common revenue model is the subscription fee that is

8 McKenna et al., "Smart Meter Data: Balancing Consumer Privacy Concerns with Legitimate Applications," Energy Policy 41 (2012): 808.

9 McAfee and Brynjolfsson, "Big Data: The Management Revolution," (2012): 4.

10 Muschalle et al., "Pricing Approaches for Data Markets," in International Workshop on Business Intelligence for the Real-Time Enterprise, Springer (2012): 130.

11 Schomm et al., "Marketplaces for Data: An Initial Survey," ACM SIGMOD Record 42, no. 1 (2013): 16.

12 Ibid.: 19; Hartmann et al., "Big Data for Big Business? A Taxonomy of Data-Driven Business Models Used by Start-up Firms," Working paper, (2014): 11.

13 Shiller, "First Degree Price Discrimination Using Big Data," (2014): 18.

charged for the use of the service.[14] For instance, data-driven service providers could use internal and external data stocks to optimize advertising services, using internal and external data stocks to improve credit check services. Thirdly, data becomes a tradable good where data stocks are licensed or sold to others. [15] For instance, map data could be temporarily accessed for a defined period of time.

The special characteristics of big data[16] to be of *high*-volume, *high*-velocity and *high*-variety have the negative side-effects for consumers that privacy regulation is getting more complex. Moreover, there is a lack of awareness fostered by the characteristic data broker industry[17] and the common features of the Internet of Things technologies:[18]

1. *Embeddedness:* Embedded sensors automatically and massively collect personal data which is largely hidden from users' awareness.
2. *Profiling:* The potentials for deriving information and detecting habits from data is generally non-transparent to users
3. *Default:* Measurability is becoming the default case and is increasingly hard to avoid.

The traditional answer of this obsessive collection of data is the principle of data minimisation and providing privacy by design mechanisms.[19] In general, this is useful whenever it can be applied. In many cases – even from a consumer perspective – this answer is too short-sighted as it neglects the trade-offs between risks, but also consumers' benefits of data services.[20]

For instance, the new possibility of Smart Meters does not just allow e. g. utility providers to improve their Smart Grid management – which was one argument for the nationwide roll-out – but also provide new means to improve the energy literacy of consumers and their knowledge about their consumption behavior.[21] However, it can be shown that the most attractive added values for the users need a real-time, fine-grained and long-lasting collection of the electricity consumption. Regrettably, expansive data collection has not only the highest value for the consumer, but also the highest potential for abuse. Still, a single-sided view on

14 Hartmann et al., "Big Data for Big Business? A Taxonomy of Data-Driven Business Models Used by Start-up Firms," (2014): 10.
15 Ibid.
16 McAfee and Brynjolfsson, "Big Data: The Management Revolution," (2012): 4.
17 Federal Trade Commission, "Data Brokers: A Call for Transparency and Accountability," Washington DC: Federal Trade Commission, (2014): 46.
18 Atzori et al., "The Internet of Things: A Survey," Computer Networks 54, no. 15 (2010): 2789.
19 Pfitzmann and Hansen, "A Terminology for Talking about Privacy by Data Minimization" Technical Report. TU Dresden and ULD Kiel, (2010): 6.
20 SVRW, "Digitale Souveraenitaet," Gutachten des Sachverständigenrats für Verbraucherfragen (2017), 1, https://www.bmjv.de/SharedDocs/Downloads/DE/PDF/SVRV_Gutachten_Digitale Souveraenitaet.pdf?__blob=publicationFile&v=2. [Access date: 10/08/17].
21 Schwartz et al., "What People Do with Consumption Feedback: A Long-Term Living Lab Study of a Home Energy Management System," Interacting with Computers 27, no. 6 (2015): 554.

technical security and neglecting user's demands on data-driven services endanger not achieve the broad market, but also user acceptance.[22]

For this reason, there are discourses in consumer research at the moment how to develop the principle of data minimization towards the guiding principle of digital sovereignty. The latter encompasses three cornerstones: consumers' digital competence, consumer-friendly regulation, and consumer-friendly technical measures.[23] We want to contribute to this discourse by outlining the different dimensions of personal data playing a role in upcoming data economies and outline personal and market mechanisms that have an effect on data use and protection.

2. Dimensions of personal data

What privacy is and what personal data are cannot be answered easily. For instance, in early days of the discourse privacy was proclaimed as the right to be let alone;[24] with the rise of information technology privacy was understood as the claim to determine for oneself when, how and to what extent information is communicated to others;[25] and with the growth of the Internet privacy is discussed as contextual integrity.[26] So, instead of giving a final definition, technological advances should always give reasons to rethink established concepts and definitions in the light of how new possibilities will have and could have an impact on social practices.[27]

With regard to the new possibilities and the resulting new complexity of collecting, storing and using big data we want to discuss personal data with regard to the following three dimensions:

a) Personal data as intimacy

The first and the most common definition of personal data understood it as any information that is connected to a person such as the personal eye color, the personal income, etc. and that this information could be used to identify a person such as the phone number or IP address of a computer client.

The trading of such kind of personal data must respect the strict data protection regulations. Therefore, technical methods for anonymization and pseudonymization are often used in order to prevent an identifiability of individual

22 Stevens et al., "Mehrseitige, Barrierefreie Sicherheit Intelligenter Messsysteme" Datenschutz und Datensicherheit (DuD) 38, no. 8 (2014): 538.
23 SVRW, "Digitale Souveraenitaet," (2017): 5.
24 Warren and Brandeis, "The Right to Privacy," Harvard Law Review, (1890): 195.
25 Westin, Privacy and Freedom, (1968): 7.
26 Nissenbaum, "Privacy as Contextual Integrity," Wash. L. Rev. 79 (2004): 120.
27 Solove, "Conceptualizing Privacy," California Law Review, (2002): 1129.

persons from data stocks, to be able to market the data more easily.[28] The challenge, however, is that undetected patterns may be found in big data stocks, from which individual persons can be assigned with a relatively high probability because of data mining algorithms and the combination with other data stocks.[29] This is one of the reasons why trading raw data is less common today than providing data-driven services, because this allows data agents a better control about the specific use of the data.

In Internet of Things era, this view does not become obsolete. Still, it does not answer several issues dealing with personal data, e.g. that keeping data secret may not be the primary concern of users, but they want a reward for sharing personal data on the one hand and to avoid negative impact on the other.

b) Personal data as consumers' work effort

The idea of getting a reward in return for data is closely linked with the idea that personal data are the outcome of consumers' work effort. Regarding this, an analogy between the work of consumers and the one of media professionals exists. In particular, the digitalization has democratized the production and distribution meaning that nowadays everyone could easily write texts, take pictures, record videos, create songs and publish their (master)piece of work.[30] This had led to a debate about fair pricing models, new copyright models, etc. using, adapting and sharing such kind of user generated data.[31] With the rise of digital sensors, it is likely that a new discussion about ownership and fair pricing of sensor-generated content emerges as users, sensors, and algorithms are inseparably mixed together in the creation of value of such data so that an attribution of the individual contribution will be difficult.

There is another relation between artistic achievements and data collected by consumers' devices regarding the twofold character of self-expressions. For instance, a self-portrait is not just a result of the production work, but expresses the individual style, taste, and personality of the artist, too. To a certain respect, a data stock presents a self-expression of the consumer as it tells a lot about the consumer's personality and life-style. However, the main difference is that in the one case it is an intended self-expression, while in the other case it is typically an unintended one. In general, we could clearly define what user generated data and what machine generated data is. However, on a technical-syntactical level such

28 Pfitzmann and Hansen, "A Terminology for Talking about Privacy by Data Minimization" (2010): 9.

29 Witten et al., "Data Mining: Practical Machine Learning Tools and Techniques" (2016): 36.

30 Bauer, "User Generated Content–Urheberrechtliche Zulässigkeit Nutzergenerierter Medieninhalte," in Nutzergenerierte Inhalte Als Gegenstand Des Privatrechts (2010): 6.

31 Goss, "Codifying a Commons: Copyright, Copyleft, and the Creative Commons Project," Chi.-Kent L. Rev. 82 (2007): 963.

distinction is far less clear. In particular, one cannot see by looking at the data, if they resulted from intended action or an unintended side effect, e. g. using a consumer device. For instance, somebody is driving down a street, and the sensors installed in the car detect a free parking space. This information together with the location of the car is send to a cloud services.

From a technical point of view, it makes no difference whether the driver has driven through the city in order to discover free parking places, or whether the free parking area has been accidentally discovered, e. g. while the driver was on the way to work. This raises the question, if e. g. users could prohibit data transfer not just because of privacy reasons, but also because to have an exclusive right to make use of the data – and the other way around that e. g. car manufactures could prohibit not turning off the data transfer because the data collection is part of their business model.

c) Personal data as a cause for negative impact

Pragmatists like Solove suggest to understand privacy protection as a strategy to guard against disruptions to certain practices.[32] This view defines personal data not in the traditional way as *data sources* that *relates* to a person, but as *data usage* that has an *impact* to a person. This is a slightly, but significant difference.

This issue could be illustrated e. g. by the upcoming Connected Car technology. As the car sensors observe the driving behavior, the collected data are related to the individual driver – and hence it is personal data in the traditional sense outlined above. Car manufactures might use the aggregated data to calculate prices on a region or country level. In this case, the data is statistically anonymized, so, regarding to the traditional definition, it is no personal data anymore. However, data usage still has an impact on the individual. Hence, the consumers might get informed about such kind of data usages, even if it falls under the apparently inoffensive category of so-called *anonymous data usage*. This applies even more since the used data was generated by the driver through using the car as outlined above.

Concerning the outlined principle of fair pricing of user-generated data, consumers should have to be informed about such (indirect) negative impacts of the information disclosure about their driving behavior. Otherwise, it has to be proofed to what extent liability claims might arise from the negative impacts using personal data. In addition, consumers should not have ex-ante control mechanisms when the data was collected, but ex-post control mechanisms when the data was used.[33]

---

32 Solove, "Conceptualizing Privacy," (2002): 1129.

33 Stevens and Wulf, "A New Dimension in Access Control: Studying Maintenance Engineering across Organizational Boundaries," in Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work (2002): 200.

In summary, all three categories of personal data are relevant dealing with data economies on an analytical and a practical level, too. The trend towards user- and sensor-generated data makes it more and more difficult to distinguish between data that – directly or indirectly – is a) generated by individuals, b) related to individuals, and c) has an impact on individuals. Hence instead to consider the categories as mutually exclusive, they are more complementary, partially conflicting views from with the data utilization should be studied and regulated.

3. Personal and market mechanisms of privacy regulation

Within data economies, users play a key role, both as supplier as well as target of data analysis. Hence, the question arises, what users' motivation to protect or disclose personal data is.

A well-known explanation in literature is the one of the privacy economy model proposing that the user is a 'rational actor' who determine benefits and costs in privacy decisions.[34] Dinev and Hart assume that the balancing of risks and benefits rely on a mental privacy calculus.[35] From this stance, several laboratory experiments were conducted to elicit the implicit monetary value of personal data. For instance, Feri et al. investigated in an experiment the willingness to publish the results of an intelligence test online for a monetary reward.[36] The results show that the participants performing below-average were less willing to expose their data. Jentzsch conducted a similar study where participants had the option to sell their private data over a second-price auction and explored how participants assessed privacy and reputation.[37] Grossklags and Acquisti investigated the (un)willingness to sell personal information by asking e. g. the maximum willingness-to-protect for personal data.[38] Beresford et al. analyzed the (un)willingness to pay more for a good in an online store, if less personal information was requested.[39]

In summary, these experiments indicate that users are not willing to pay much money for improved data protection. However, these findings must be handled with care because of the limited ecological validity. In addition, even consumers

34 Acquisti et al., "The Economics of Privacy," Journal of Economic Literature 54, no. 2 (2016): 443.

35 Dinev and Hart, "An Extended Privacy Calculus Model for E-Commerce Transactions," Information Systems Research 17, no. 1 (2006): 62.

36 Feri et al., "Disclosure of Personal Information under Risk of Privacy Shocks," Journal of Economic Behavior & Organization 123 (2016): 140.

37 Jentzsch, "Auctioning Privacy-Sensitive Goods:" in Privacy Technologies and Policy: Second Annual Privacy Forum, APF 2014, Athens, Greece, May 20–21, 2014. Proceedings, ed. Bart Preneel and Demosthenes Ikonomou (2014): 137.

38 Grossklags and Acquisti, "When 25 Cents Is Too Much: An Experiment on Willingness-To-Sell and Willingness-To-Protect Personal Information." in WEIS, (2007): 6.

39 Beresford et al., "Unwillingness to Pay for Privacy: A Field Experiment," Economics Letters 117, no. 1 (2012): 25.

have an increasing awareness that personal information is utilized by online providers, the supply of data seems to be more a side effect of using services rather being conscious about trading data in exchange for services. From this stance, the phrase that users pay with their data is at least as misleading as it implies that users make deliberate, conscious, well-informed decisions.

In addition, the privacy economy research relies on a methodological individualism that tries to explain the social phenomena of privacy from the motivations and actions of individual agents. By its very nature, such methodology, however, neglects market-related mechanisms of privacy regulation, for instance the (un)willingness of data agents to protect the privacy of users as their data suppliers.

One reason for data agents to respect the privacy of clients is, of course, to be compliant with data protection laws. However, with regard to market-related consumer policy economic reasons for privacy regulation should be taken into account as well. Such economic reasons become more evident when we analyse data-economies in the light of multi-sided markets.

In general, multi-sided markets are described by three characteristics. Firstly, it is important to have at least two independent but complementary user groups. Secondly, the benefit of one group changes with the size of the other group. Thirdly, an intermediary is needed which connects both groups. [40] A perfect competition on multi-sided markets works only if all sides benefit from this model and no single side is advantaged. Otherwise, the other parties will move to another multi-sided market with better conditions. In this respect, the market mechanisms themselves represent a kind of "invisible hand"[41] that keep a balance between all market actors, as this is essential for market success.

Data-driven business models of companies like Facebook, Google and others could be modelled as multi-sided markets where users taking the role of data providers, the service providers taking the role of market operators and advertisers taking the role of data clients. The outlined mechanism of multi-sided markets suggests that the data-market operator will respect a minimum level of privacy protection, fair pricing, and preventing adverse impacts. The fulfillment of these criteria is essential for the success of companies, even if laws would not enforce it. If a market operator would misuse users' personal data, they would lose customers rapidly with negative network effects as a result. Therefore, they have a self-interest in using generated private data wisely for services with an added-value for the market actors to keep balance between the market sides.

However, we never have a perfect competition, but must cope with the danger of a market failure. Digital economies are particularly vulnerable because of the

---

40 Clement and Schreiber, *Internet-Ökonomie*, 3. Aufl. (2016): 266.

41 Smith, 'Invisible Hand,' in L. S. Stepelevich, ed., The Capitalist Reader, New York: Arlington House Publishers, (1977): 20.

non-transparency of the data processing leading to information asymmetries.[42] In addition, there is a natural tendency towards winner-takes-it-all effects. This tendency is characterized by short competition periods where one competitor dominates because of scale, network and lock-in effects with the result that most other suppliers have been pushed out of the market in the long run. [43]

Such monopolistic structures can be considered as harmful because market operators become less dependent on respecting the interests of users as their data suppliers. This raises new demands for the regulation of such markets to prevent monopolies. A common strategy is to ensure interoperability to reduce lock-in and network effects. However, the devil is in the detail as technical and economic issues are typically intertwined. For instance, specifying and regulating standardized data formats and open interfaces present a key instrument to ensure the free flow of data. It is important to create a transparency in algorithms and make them verifiable as well as develop standards for interoperability between platforms.[44]

III. Conclusion

In the last decades, we observe an ongoing exponential growth of the computer power that has changed our societies. The technological progress and the huge computational capacities available today drive actual digitalization trends such as Artificial Intelligence, Big Data and the Internet of Things.

The fuzziness and complexity of the collected data on a massive scale raises new challenges to protect consumers' interests concerning privacy, fair remuneration and the offset negative impact on personal life. This asks for effective mechanisms supporting the consumers' digital sovereignty, which on the one hand acknowledges the fact that one does not know the exact use during data collection and on the other hand, protects the individual as well as the society from negative consequences. While we could not give a final answer, we would like to conclude with a brief outlook that indicates the direction of how a consumer-friendly development might look like.

Concerning the technical level, the principle of data minimization and enforcement of privacy by design are not obsolete concepts, but should be applied whenever possible. However, they should be supplemented by concepts that could be called traceability, transparency and accountability by design. The aim is to support multi-sided security by informing consumers (or consumer agencies in behalf) about what data was collected, processed, and used for what purpose. This demand on technical infrastructures becomes more relevant for a

42 Akerlof, "The Market For lemons: Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, (1970): 490.
43 Clement and Schreiber, *Internet-Ökonomie*, 3. Aufl. (2016): 218.
44 SVRW, "Digitale Souveraenitaet," (2017): 5.

distributed data process where various algorithms and data stocks are involved. In such a scenario, technically supported traceability is crucial to uncover who is responsible in the event of damage or loss.

Concerning the individual level, established concepts like the informed consensus or the earmarked use of data are also not obsolete, but should be supplemented by measures that support the digital competence of consumers. For instance, data-driven companies maybe ought to place a warning notice on their products that the use of their products could have a negative impact – for instance that an innocuous post might cause a dismissal or that a foreign country might refuse the border crossing. Another topic will be the users' power of choice about the things around him as well as the question of whether there is a right of unobserved use.

Concerning the market level, existing regulations on data protection should be supplemented on regulations that promote the diversity and competition within data markets. In well-functioning data markets, data agents must respect the interests of consumers as data providers at a minimum level. Hence, measures like ensuring interoperability and a free flow of data are welcome to increase competition by reducing lock-in and network effects. In particular, it must be ensured that data companies do not become system relevant and be too big to fail. Otherwise, competition as well as consumer friendly regulations could not be enforced.

Last but not least, digital societies and economies more and more rely on digital infrastructures constituted by big data stocks and the procedures to processing them. Because of this, the transparency, accountability and the control of these critical infrastructures must be ensured.

Zusammenfassung

Die technologische Entwicklung des Computers und neue Optionen zum Sammeln, Speichern und Übertragen von Daten in großen Mengen haben in den letzten 40 Jahren die Welt verändert. Darüber hinaus werden die Etablierung des Internets als kritische Infrastruktur und die Möglichkeiten der ubiquitären Sensorsysteme aufgrund des laufenden Fortschritts der Rechenleistung einen dramatischen Einfluss auf die Volkswirtschaften und Gesellschaften in der Zukunft haben. Wir geben einen kurzen Überblick über die technologischen Grundlagen, insbesondere im Hinblick auf das exponentielle Wachstum von Big Data und die aktuelle Umstellung auf sensorgestützte Datenerfassung. Vor diesem Hintergrund überdenken wir die verschiedenen Dimensionen von personenbezogenen Daten und Marktmechanismen, die Auswirkungen auf Datenverwendung und -schutz haben.