Sarah Spiekermann-Hoff and Rainer Böhme and Alessandro Acquisti and Kai-Lung Hui

The Challenges of Personal Data Markets and Privacy

Article (Accepted for Publication)

Original Citation:

This version is available at: http://epub.wu.ac.at/5496/
Available in ePub^WU: March 2017

ePub^WU, the institutional repository of the WU Vienna University of Economics and Business, is provided by the University Library and the IT-Services. The aim is to enable open access to the scholarly output of the WU.

This document is the version accepted for publication and — in case of peer review — incorporates referee comments.
The challenges of personal data markets and privacy

Sarah Spiekermann 1,*

Phone + 43 1 313 36 5460
Email sspieker@wu.ac.at

Alessandro Acquisti †

Rainer Böhme †

Kai-Lung Hui †

1 Vienna University of Economics and Business (WU Vienna), Welthandelsplatz 1, D2, 1020 Vienna, Austria

Abstract

Personal data is increasingly conceived as a tradable asset. Markets for personal information are emerging and new ways of valuating individuals’ data are being proposed. At the same time, legal obligations over protection of personal data and individuals’ concerns over its privacy persist. This article outlines some of the economic, technical, social, and ethical issues associated with personal data markets, focusing on the privacy challenges they raise.

Keywords

Responsible Editors: Rainer Alt and Hans-Dieter Zimmermann
Why personal data markets matter

Personal data is the new oil of the Internet and the currency of the digital world.

In 2009, Meglena Kuneva, speaking as European Commissioner for Consumer Protection, compared personal data to oil, in order to illustrate how information pertaining to individuals had become a crucial asset in the digital economy. Every day, individuals around the world send or receive 196 billion e-mails (Radicati 2014), submit over 500 million tweets on Twitter (Internet Life Stats 2014), and share 4.75 billion pieces of content on Facebook (Noyes 2014). A report by the Boston Consulting Group (2012) projects that the sectors leveraging personal data will leap ahead of the rest of the economy and produce €1 trillion in corporate profits in Europe by 2020.

The World Economic Forum (2011, 2012) has described personal data as a new asset class, and a complex ecosystem of entities collecting, analyzing, and trading personal information (companies such as BlueKai, Avarto, Rapleaf, Accurint, and Merlin) has emerged. Personal data is seen as a new asset because of its potential for creating added value for companies and consumers, and for its ability to enable services hardly imaginable without it. Companies use personal data for a variety of purposes: reduce search costs for products via personalized and collaborative filtering of offerings; lower transaction costs for themselves and for consumers; conduct risk analysis on customers; and increase advertising returns through better targeting of advertisements. Personal data can also be a product in itself, when it is entangled with user-generated content, as in the case of social media. Personal data can also become strategic capital that allows businesses to derive superior market intelligence or improve existing operations. This can materialize in better or new forms of product development (for instance, mass customization: Henkel and von Hippel 2005) as well as price discrimination (Acquisti and Varian 2005). Businesses can also build competitive advantage or create market entry barriers by using personal information to lock customers in (Shapiro and Varian 1998).
At the same time, personal data can become a burden for organizations as much as an asset. One of the most salient liabilities of holding personal data arises from the legal uncertainty surrounding its management. Privacy regulation, which comprises the protection of personal data, is an evolving and among the least globally harmonized fields of law. Different countries in the world use different definitions of personal data and apply different rules governing its collection and use. As a result, businesses operating in a digital economy without borders are exposed to legal and enforcement risks (Romanosky et al. 2012) that are hard to quantify. Other liabilities arise from the risk that large collections of personal data become targets of cybercrime, in particular when they include identifying or financial information. To mitigate this threat, companies must exert constant effort, adjusting protection technology and organizational processes to protect information assets and secure data exchanges. But even when they do so, a state of zero risk remains unachievable. Organizations that strive for compliance can still fall victim to data breaches, which in most jurisdictions entail costly breach notifications that damage a firm’s reputation and market value (Acquisti et al. 2006). The extent of ex post loss incurred through such breaches is hardly foreseeable. Its extent could be huge, and yet not fully traceable. For example, the damage caused by the data breach incident in Sony’s PlayStation Network in 2011 was estimated alternatively at €128 million, and at €3.5 billion.¹ A major reason for the unpredictability of losses is the unforeseeable shift in public opinion’s responses to personal data breaches. Public opinion can turn against a firm considered responsible for its breach, damaging its brand or even a whole industry. Breaches can trigger rash government intervention. But they can also go largely unnoticed. Adding to this tension is, since the Snowden revelations, the deep conflict between law enforcers’ desire to tap personal data and a sensitized media to this issue.

Against the background of these market promises as well as economic, social and political risks, we aim to offer an academic perspective on personal data markets. We discuss where these markets stand, legally, technically and ethically, and highlight the major questions that market players and policy makers will arguably need to face in handling those markets.

**Personal data markets and their challenges**
Researchers anticipated personal data markets since the early 1990s (Laudon 1996). When the idea was first proposed, by and large the many members of the academic legal community reacted cautiously to it, under the argument that people’s privacy concerns and their legal right to privacy would challenge the treatment of personal data as an asset class (Acquisti and Varian 2005). To many academics, it seemed as if the ethical right to privacy would be antithetical to the very idea of markets for personal data (Samuelson 2000).

While many still consider privacy an inalienable human right, and while privacy-enhancing technologies have been designed to protect it, data markets have developed in the opposite direction. Due to Internet users’ apparent comfort with sharing their data, more and more organizations today engage in the trading of consumer data, operating in legal grey zones when it comes to handling personal information assets.

Discussions around more legitimate personal data markets have recently emerged, pushed by influential global actors like the World Economic Forum. In the US and in Europe, it is now intensively discussed whether personal data could be seen as “property” (Schwartz 2004; Purtova 2012; Spiekermann and Novotny 2015). The OECD and scholars around the world have started to think about how personal information could be priced (OECD 2013; Acquisti 2010, 2014; Spiekermann et al. 2012). Thought leaders have proposed whole new market structures and business models that may allow consumers to get into the driver’s seat for their personal data (Searls 2012; Hamlin 2013). Established software vendors envision the technical architecture that would be needed to do so. And start-ups such as Qiy, Connect or Gigya go ahead and propose tools and services that they believe make it happen.

Against these developments towards personal data markets, however, also economic, legal, technical and social challenges have emerged. We discuss some of those challenges in the rest of this section.

Intertwined legal and economic challenges

In many countries, the use of personal data is highly regulated, as is the data exchange between jurisdictions. Most often, a set of data protection
principles that inform privacy laws include rights and obligations such as data minimization, legitimate use, purpose binding, and informed consent (see, for instance, the OECD Privacy Guidelines). These principles leave little room for market negotiations between the data subject and the data controller, let alone between third parties. Thus, personal data markets must deal with these constraints or operate in grey areas—as many currently do. For example, some firms use enforcement gaps or regulatory arbitrage between jurisdictions to engage in the trade of personal data. As a result, aside from special and tightly regulated ventures (such as address brokerage for direct marketing, credit reporting, government health insurance, or pollster panels), most of personal data gathered today online seems to remain a rather inconvenient unit of account.

If that was to change, and a true kind of “currency” or monetary value were to be sought for personal data, then we would need to deal with the fact that data has—in many respects—the traits of a free commons. By its nature, personal data is non-rival, cheap to produce, cheap to copy, and cheap to transmit. It is substantially different from typical commodities, such as oil. Hence, markets for personal data would need to rely on legal frameworks that establish alienability, rivalry, and excludability for personal data, and assign initial ownership to an entity such as the data subject. To enforce asset rights, the right institutions, sanctions, and—most importantly—technology must interact seamlessly.

The possible economic consequences of a property right regime for personal data may include scarcity of personal information and competition in its use. Data controllers with data-intensive business models may need to adequately compensate data subjects, whereas business models no longer profitable when personal data is costly may vanish or pursue other sources of revenue, such as service fees. Will market forces thus lead to “data rationalization,” a kind of profit-motivated data avoidance? To answer this question, one would need to study the shifts in market structure and the industrial organization of the affected sectors (Goldfarb and Tucker 2011). If corporate information systems need upgrades to support enforcement of property rights, businesses, hardware supply chains, software development, and vendors will be affected. An economy’s potential to innovate may be affected as well.

Another economic challenge associated with personal data markets relates
to how to value personal information. Due to context-dependence and contingencies that affect the costs and benefits arising from the protection or the sharing of personal information, evaluating personal data is notoriously difficult (Berthold and Böhme 2009). Numerous experimental economic studies have attempted to capture price tags for personal data items (Grossklags and Acquisti 2007; Lesk 2012; Jentzsch et al. 2012), but it is hard to map these data points into a consistent picture, due to heuristics and biases which may significantly affect individuals privacy choices, including in fact their valuations of their personal data (Acquisti 2004; Acquisti et al. 2013). This task becomes even more difficult if one tries to crosscheck against prices in advertising (NA 2013) or credit markets (Böhme and Pötzsch 2010).

The difficulty of measuring the value of data raises many questions about price discovery in personal data markets. How can buyers and sellers negotiate in a setting where information is inherently asymmetric? What market mechanism can determine the right price under the constraint of minimal information leakage? How should buyers and sellers do the accounting for their trades? And how can auditors detect fraud or bust cartels? On the fiscal side, how should data trading or income from the sale of personal data be taxed? On a macroeconomic level, how will personal data trade affect gross value added and the balance of payments? Obviously, the answer to these questions is not independent of technology. For example, noisy, pseudonymous, or anonymous data may be less valuable than fully identifiable records, but by how much? And will the difference match the data subject’s preference for being less identifiable? Even if stable prices for personal data were found and personal data markets were liquid, there would be no guarantee that prices will reflect any “intrinsic” value of data. History is full of periods of mispricing, some of them lasting decades, in virtually all asset markets; why should personal data markets be exempt? But then, how can we proactively avoid speculative bubbles, and how can we ensure that personal data markets will not detach from the real lives of people?

Beyond value and price, a more general concern is how the mere existence of personal data markets may affect society. Consider, for instance, strategic data subjects who maximize the value of their personal data and therefore engage in strategic behavior, such as avoiding leaving traces that link them to people from troubled neighborhoods. How does this affect
social cohesion, equality of opportunity, freedom, and democracy?

Furthermore, future uses of data should be anticipated and accounted for in the pricing mechanism in order to attain fair valuations of personal data. This, however, is exceedingly difficult to achieve for information shared online and in particular via online social networks, where users often cross-post information about others and so impose privacy externalities on each other (Tufekci 2008; Biezok and Chia 2013; Cao et al. 2014). The challenge is how to internalize such privacy externalities so that the data subject can be fairly compensated. The standard solution is to facilitate bargaining among the data subjects and data controllers (Coase 1960), but this goal may prove nearly impossible to attain when the community has millions of users. The other solution is to restrict the sharing and use of personal data, but such policy can distort service quality and social welfare (Cao et al. 2014).

Finally, if personal data becomes property, important legal challenges will include tailoring property rights so that they are compatible with the notion of privacy as a fundamental right, defining the initial allocation of property rights, balancing sanctions, and seeking international coordination. Tailoring rights means restricting alienability and exclusivity. For example, people should always retain the right to use their personal data in private contexts, and the government should have free access to one’s date of birth for purposes such as identification. Defining the initial allocation is difficult when a data item concerns the relation between multiple data subjects, such as links in a social graph. Sanctions must be enforceable and technology-neutral. Specific to personal data is the problem of false information and the risk of false accusation, which bound the toughness of sanctions. Given that it took decades to negotiate trade agreements for (selected) tangible goods or to find a common understanding on intellectual property rights, the size of the challenge to harmonize property rights for personal data is hard to gauge.

Technical challenges

Arguably, the only way to enforce property rights for personal data would be by mandating technology that “reverses” the laws of information goods. Scholars have pointed out that this endeavor has parallels with digital rights management (DRM), which protects media content (Kenny and
Korba 2002), but companies are still reluctant to adopt this technology for privacy, mainly due to an overt lack of incentives (Böhme and Koble 2007).

Solutions originally developed as privacy-enhancing technology may be applied to personal data markets. Methods invented for the privacy principle of data minimization may for instance be employed in systems to enforce fine-granular data access rights. These rights could be encoded in new formal languages, drawing on languages specified for privacy policies, such as P3P (Cranor 2002) and EPAL (Schunter and Powers 2003), and attached to data as sticky policies (Pearson and Mont 2011). For example, cryptographic zero-knowledge proofs, running on top of an anonymous communications layer, can help to exchange only the necessary information with a transaction partner (Chaum 1985). The effectiveness of two anonymous credential systems, IBM’s Idemix and Microsoft’s U-Prove, has been demonstrated in pilot studies (ABC4Trust, PrimeLife). Fully homomorphic encryption is on the horizon (Gentry 2010), but it is still too inefficient for bulk processing of personal data. As a result, an existing challenge is that cryptographic protocols require all transaction details to be known at the time of invocation. This requirement is contrary to the needs of personal data markets. Many features that make personal data particularly valuable require flexibility during a transaction or completely dissolve the association between data and transaction. For example, it could be argued that data stimulates innovation when it is shared without knowing in advance what others will figure out to do with it.

A promising workaround is to relax the trust assumption: instead of using cryptography to secure the data itself, cryptography could secure the integrity of a trusted computing platform and trusted software, which runs conventional algorithms on unencrypted data. This usage control paradigm—as opposed to access control—trades technical complexity for new organizational challenges including institutional trust relationships, supply chain security, interfaces and processes for effective audits, and sanctions to deter abuse (Sackmann et al. 2006). Systems enabling a flow control for accountable information would have to be designed, implemented, tested, and standardized along with the underlying data models, audit processes, and controls. Quality metrics could adapt existing paradigms for privacy metrics, such as $k$-anonymity for database privacy (Sweeney 2002),
effectiveness of data perturbation approaches (Menon et al. 2005; Li and Sarkar 2006), information-theoretic metrics of data leakage, or randomization-based proof techniques inspired by differential privacy (Dwork 2011).

Such a transformation of enterprise systems would need to go along with novel end user devices. Here, one can envisage interfaces that empower individuals to manage their personal data (Zwick and Dholakia 2004), ideally on their own trusted devices; this practice would establish a strong base for the user’s digital identity. Much has been proposed in this realm (e.g., Hansen et al. 2004). Personal data markets may, perhaps, create the right incentives to push adoption of such solutions above critical mass. Furthermore, substantial human factor research has improved our understanding of when and why users make privacy compromises, intentionally or accidentally. This research can feed into insight on how to build interfaces that make personal data markets usable for ordinary people.

Social and ethical challenges

Interpreting personal data as a tradable good raises ethical concerns about whether people’s lives, materialized in their data traces, should be property at all, or whether in fact personal data should be considered inalienable from data subjects. The “propertization of the human being” touches upon fundamental discourses in philosophy, sociology, and political sciences about what is private and public, what constitutes identity and what it takes to be a responsible (in German: “mündiger”) citizen with sufficient liberty to form preferences and opinions by him or herself. Disagreement between cultures and schools of thought on these issues is hard to reconcile. For centuries, philosophers have tackled the question of identity. But the joint effect of technological advances and market forces renders people into “data subjects” whose “digital identities” are traded and used (potentially without their knowledge and consent), and forces age-old thought experiments to confront a new reality. Moreover, a digitally networked society without borders urges us to find global consensus on these issues as it becomes much harder to preserve islands of idiosyncrasies.

Researchers in the information systems community, in legal studies, in philosophy, in marketing, and in decision sciences have focused their efforts on what privacy is (Solove 2005; Iachello and Hong 2007) and
how it can be measured (Smith et al. 1996; Hong and Thong 2013). User experiments have investigated why and under what circumstances people self-disclose (Dinev and Hart 2006; Berendt et al. 2005; Acquisti and Grossklags 2005) and what decision-theoretical pitfalls people regularly fall victim to when it comes to take privacy decisions (John et al. 2011). Based on such insights, measures to “nudge” people into what is considered more rational privacy behavior have been investigated (Acquisti 2009). The common social research perspective to protect privacy is underlined by Daniel Solove’s (2005) overview of the privacy construct. Solove’s legal analysis of what Western cultures consider to be “privacy” reveals that the commercial data handling practices now promoted by personal data market proponents seem to completely undermine or even dissolve this value. Common market practices, such as the aggregation of personal data, identification, secondary use, exclusion, and decisional interference are all recognized privacy breaches according to Solove’s taxonomy. This seems to be a dilemma in which personal data markets operate for now.

Embracing market-based frameworks, some early research has started to look into stock market reactions to personal data breaches (Acquisti et al. 2006) as well as pay-for-privacy schemes (Jentzsch et al. 2012). Microeconomic modeling and reflections on personal data’s role in markets have been provided (Hui and Png 2006). A few insights exist into how people value their personal data (Spiekermann et al. 2012; Huberman et al. 2005; Danezis et al. 2005; Hann et al. 2007; Hui et al. 2007; Grossklags and Acquisti 2007; Acquisti et al. 2013). But can personal data really evolve to become an asset for people? Can people develop a psychology of ownership for their data in the same way as they do for tangible assets? Will people not want to continue freely communicate online, chat, talk, post and provide their data? Pricing psychology for the intangible data asset is an open research field. At the same time, people are heterogeneous in terms of privacy preferences and hence their potential participation in markets for personal information (Berendt et al. 2005; Awad and Krishnan 2006; Norberg et al. 2007). The question is therefore whether the desire to stay private may disadvantage some people in personal data markets more than others. And if we assume that individuals trade personal data, what kind of controls and guarantees do they want and need to trust in the market they participate in?
A final outlook on personal data markets

The phenomenon of personal data markets is gaining momentum. It is strongly promoted by industry players that benefit from trade in personal information. These are not necessarily traditional companies in electronic markets investing in customer relationships and directly offering valued products and services to consumers. Instead these “first parties” have in recent years given up an important part of the data market to “third parties” who collect, aggregate, infer, resell and package users’ data. Mostly, we would argue, doing so without any ordinary user expecting that this is happening.

The problem we see is that most users today only know about and (sometimes) trust those ‘first parties’ they see and interact with. If they learned about today’s volume and business done with their data among third parties, they may be surprised and feel betrayed. No matter whether and to what extent first party companies have engaged in data deals themselves, they could all be hit by a backlash from users once they find out. A general air of mistrust would be the result in which data-intensive industries will have difficulties to innovate and would need to heavily invest into regaining trust.

Many recent studies (including the study contributed by BCG staff in this special issue) show that people only accept active data sharing where they are consciously involved in the data exchange. They don’t appreciate passive data collection. Third-party use of data is seen rather negatively, in an identified as well as anonymous form. What seems acceptable to consumers are situations where the companies they deal with process their data for the respective service relationship and guard contextual integrity. When these first parties want to use the data further they need to offer customers appropriate returns. But they cannot expect that everyone will agree to such secondary uses and hence they need to give their customers a true choice over participation (Roebert et al. 2015).

Our position—against this background—is that companies, which hold customer relationships should go back to more trustworthy relationships with their customers. This implies that they need to respect peoples’ data protection expectations and consider much more carefully how they engage with third parties. First parties should not rely on ‘data-deals’ too much,
but compete on service and product quality. They should give customers the option to pay for online services that are fully privacy preserving and only allow for data sharing with third parties if customers allow this to happen and get a fair share of the deal in a transparent way. If it is impractical to explicitly spell out future sharing of data with third parties and secondary use, they should communicate the situation transparently with consumers and let the “price” of such sharing/use to emerge through an efficient pricing mechanism in the market. Where data is shared and used, it should be ensured that the terms of the deal agreed on with the data subjects are technically and legally respected (see, e.g., a technical proposal for this by researchers from Microsoft in this special issue: Maguire et al. 2015).

We believe that if companies move into this kind of transparent business scenario many legal challenges could be avoided and economic, technical and social challenges could fall into place over time.

References


AQ9

AQ10


http://eproofing.springer.com/journals/printpage.php?token=hhy8suqZZz4fePqFqhWt6HvoXQCS-GQZpOazC5dIE
Cambridge University Press.


AQ11


**CUPS CyLab Usable Privacy and Security Laboratory.**


**European Commission (2012). Proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), 25 Jan.**


AQ17

AQ18


---