myneData: Towards a Trusted and User-controlled Ecosystem for Sharing Personal Data

Roman Matzutt¹, Dirk Müllmann², Eva-Maria Zeissig³, Christiane Horst⁴, Kai Kasugai⁵, Sean Lidynia¹, Simon Wieninger⁴, Jan Henrik Ziegeldorf¹, Gerhard Gudergan⁴, Indra Speicker gen. Döhmann², Klaus Wehrle¹ and Martina Ziefle³

Abstract: Personal user data is collected and processed at large scale by a handful of big providers of Internet services. This is detrimental to users, who often do not understand the privacy implications of this data collection, as well as to small parties interested in gaining insights from this data pool, e.g., research groups or small and middle-sized enterprises. To remedy this situation, we propose a transparent and user-controlled data market in which users can directly and consensually share their personal data with interested parties for monetary compensation. We define a simple model for such an ecosystem and identify pressing challenges arising within this model with respect to the user and data processor demands, legal obligations, and technological limits. We propose myneData as a conceptual architecture for a trusted online platform to overcome these challenges. Our work provides an initial investigation of the resulting myneData ecosystem as a foundation to subsequently realize our envisioned data market via the myneData platform.

Keywords: Personal User Data, Personal Information Management, Data Protection Laws, Privacy Enhancing Technologies, Platform Design, Profiling

1 Introduction

Trading and monetizing user data is a prevalent business model for most Internet companies [CL16]. With the ongoing digitalization, more and more data becomes available. This provides even more benefits and business models for companies, e.g., in product and service optimization [CL16]. However, small and medium-sized enterprises (SMEs) have virtually no direct access to such data as the market is dominated by giants such as Google, Facebook, or Amazon [HH14]. Despite its benefits for service quality, users are concerned about a perceived loss of control over their personal data [EC11].

¹ Chair of Communication and Distributed Systems, RWTH Aachen University, Ahornstraße 55, 52074 Aachen, Germany, {matzutt, ziegeldorf, wehrle}@comsys.rwth-aachen.de.
² Goethe-University Frankfurt/Main, Chair of Public Law, Information Law, Environmental Law and Legal Theory, Theodor-Adorno-Platz 4, 60323 Frankfurt/Main, Germany, {muellmann, speicker}@jur.uni-frankfurt.de.
³ Chair of Communication Science, Human-Computer Interaction Center, RWTH Aachen University, Campus-Boulevard 57, 52074 Aachen, Germany, {zeissig, lidynia, ziefle}@comm.rwth-aachen.de.
⁴ FIR e.V. at RWTH Aachen University, Business Transformation, Campus-Boulevard, 52074 Aachen, Germany, {christiane.horst, simon.wieninger, gerhard.gudergan}@fir.rwth-aachen.de.
⁵ Formitas GmbH, Campus Boulevard 57, 52074 Aachen, Germany, kk@formitas.de

doi:10.18420/in2017_109
In this work, we strive to establish a sustainable ecosystem that brings both worlds together: We want to enable a marketplace, or data market, where users keep control over their personal user data and where they can directly share this data selectively with a large variety of data processors such as research groups or SMEs. We identify a wide field of challenges for such an ecosystem: It must address users’ privacy demands as well as legal obligations and simultaneously it must enable accurate insights for data processors. Today, providers of Internet services already collect and profit from a plethora of user data. Users tend to blindly accept this large-scale data collection to be able to use the offered service. The new European General Data Protection Regulation (GDPR), however, will thoroughly punish processors’ misconduct with severe financial penalties—up to 4% of the total annual turnover or 20 million Euro, whichever is higher (Art. 83 sec. 4–6 GDPR). This creates a pressing demand to create a lawful, secure and trusted user-data market.

As a solution, we propose myneData, a digital and trusted data market with focus on the protection of data privacy. We define an ecosystem for sharing personal user data in which users keep control over their data according to individual privacy demands and receive an (e.g., financial) compensation. Within this ecosystem, third parties request personal data directly from users giving consent to process their data. We see the great potential to show that data protection and data economy are not mutually exclusive in this environment.

2 The myneData Scenario

Figure 1 shows the scenario we assume for an ecosystem that enables the user-centered sharing of personal user data with third parties. A set of users of different online services create large amounts of manifold personal user data via multiple diverse data sources. Examples for personal user data are social media activities, sensor readings, e.g., from wearables or smart home systems, personally identifiable information (PII), or online shopping behavior. In accordance with the privacy calculus theory [DH06], we assume that users are willing to share their data as long as they remain in control and potentially are compensated. As users have individual preferences concerning privacy, each user has a sensitivity profile describing the subjective importance of their individual data sources.
This potential data pool then becomes attractive to data processors, who are interested in gaining new insights from personal user data. While providers of Internet services such as Google, Facebook, or Amazon are also data processors, small players such as researchers or SMEs can struggle to gain access to valuable data [HH14]. Those data processors thus can profit from having direct access to personal user data. They seek to obtain either anonymized datasets describing their interest group or answers to statistical queries such as average values, extrema, histograms, etc. We observe a high potential for new use cases if data processors and users can cooperate in sharing and processing personal user data without intermediate data brokers keeping the data. However, we also identify clear conflicts of interest between users and data processors. Users want to protect their privacy (according to their sensitivity profile) while data processors seek to gain as much information as possible from the personal user data. Both interests cannot be simultaneously satisfied to full extent. Thus, any design to realize a user-centered data-sharing ecosystem has an inherent trade-off between data privacy and data accuracy, where data accuracy directly influences value of the data.

In total, we identify four dimensions that influence the design space of personal-data sharing ecosystems: (i) user interests, (ii) data processor interests, (iii) legal restrictions on processing personal user data, and (iv) technical practicability of security and privacy measures needed. In the following, we will detail the particular challenges for each dimension and sketch approaches to overcome these challenges with the goal of working towards a viable realization of our envisioned myneData ecosystem.

3 Challenges for a Sustainable Data Market

In this section, we discuss challenges w.r.t. user demands, data processor needs, legal obligations, and technical possibilities arising from our scenario outlined in Section 2.

3.1 User Perspective

User demands create manifold challenges for the myneData ecosystem: provide privacy and control, usability and comprehensibility, the right compensation, and the generation of trust. Many studies show that users express high concerns regarding their information privacy on the Internet [SDX11]. Privacy is a multifaceted construct and, in the online context, the perceived control about personal information is the key aspect [We68]. Within the myneData ecosystem, the users are to be enabled to control their personal data without impairing their information privacy. The complexity of privacy protection needs to be reduced to a level that every user can understand and effortlessly manage their data within the scope of their personal needs of information privacy. With participation in the myneData ecosystem, despite feeling to be in control, personal data is distributed to even more third parties than it is already based on everyday online interactions. Thus, trust in the myneData system needs to be generated and perceived risks for privacy reduced—as well as incentives for using the system need to be given.
Trust is the willingness to be vulnerable to the actions of others based on a positive expectation [BDS10]. In December 2016, we conducted a survey with $n=200$ German participants to evaluate the ecosystem’s concept. The answers showed several perceived issues and reasons for distrust and concerns. Within the myneData ecosystem, users have to trust the service provider to keep their promises and also the third parties that request data to comply with the rules or at least the law. Additionally, they need to trust the system to be protected against intruders. Furthermore, lay persons as well as experts must be able to operate in the ecosystem. Thus, language and interface design need to be simple enough while providing sufficient information to be trustworthy and fulfill legal requirements.

The ecosystem shall provide benefits for the user to incentivize participation. In general, monetary rewards, vouchers, but also providing information and analyses, are conceivable. Several new challenges surface, as do new kinds of incentives for the user, for example, in the case of home automation and smart home systems. To also reach home automation ecosystems that are not yet connected to the Internet, it would be necessary to incentivize owners to adapt their home automation systems. The much larger and quickly growing market is smart home and the Internet of Things. Here, consumer devices are typically connected to an isolated, vendor-specific cloud system. A data market such as envisioned by myneData allows to combine these separated data pools and provide both users and data processors with convenient and combinable access to all device data.

### 3.2 Economic Perspective

The establishment of the myneData platform comes with great challenges crucial for its success. The main challenge lies within the business model itself. In order to define these challenges, it is essential to understand the concept of a platform. A platform, in the literature also described as a two- or multi-sided market, brings together two or more distinct but interdependent groups of customers [Mu15]. The incentive for these groups to join the platform lies within the value proposition of the platform business model. In comparison to other business models, a platform inherits the feature to require two different value propositions—one for each group of customers [Mu15]. Next to defining two sufficient value propositions, which hold true value to each customer group, the platform needs to establish the customer network [Jo13]. This is the main challenge for a developing platform because it is the epitome of the chicken or the egg dilemma. Since the value of a platform rises with the number of participants, the growth at the beginning is the most crucial part for a possible success. It has been stated that a platform must first gain a sufficiently large user base before other businesses can gain valuable information from joining the platform [Mu15]. Once a critical mass of users and businesses is achieved, the platform will attract more participants by direct and indirect network effects [Mu15].

To illustrate that a (logically) central platform is required to overcome economic challenges for user data utilization, we exemplarily look at the smart home sector. In contrast to, e.g., social media, this economy branch has no small set of global players (e.g., Facebook or Google). The smart home sector is much more diverse and currently relies on
selling smart devices such as motion sensors, which does not include expressing user data in monetary terms. Vendors usually rely on cloud-based control platforms, but currently no business cases leveraging the potential of the distributed sensor data pools exist. It is thus challenging to mediate between the different stakeholders, and we believe that the myneData data pool has the potential of creating new use cases in this field.

3.3 Legal Perspective

From a legal point of view, myneData scenarios hold several challenges w.r.t. to data protection, not the least from a legal regime in transition as the European and national data protection laws are undergoing extensive changes. While the ePrivacy regulation on Internet services, which will replace the ePrivacy directive, is still in a draft stage, the new GDPR has already been in effect since May 2016. Its provisions will apply after a two-year transition period starting in May 2018. To strengthen data protection, the GDPR introduces several instruments which were so far unknown to European and national law, e.g., the right to data portability (Art. 20 GDPR), the requirements of data protection by design and by default (Art. 25 GDPR), or the process of a data protection impact assessment for risky data processing (Art. 35 GDPR). Furthermore, it contains many flexibility and regulatory specification clauses allowing and obliging the European member states to enact deviating rules. Consequently, there will still not be a completely uniform data protection law within the EU. Moreover, the existing national data protection laws are also facing fundamental reforms to fit the requirements and use the possibilities of the new European framework.

Many details of the new instruments introduced by the GDPR are still unknown. A scientific and pan-European debate about the framework is still in its early stages and final rulings on disputed questions, which can only be made by the European Court of Justice, cannot be expected in the near future [Mü17]. Thus, services such as myneData have to find solutions under these legal difficulties.

With regard to myneData, additional challenges arise from the different sources or situations of data collection and the broad variety of categories of data processed. Depending on the specific form and circumstances, different laws with specific requirements for the processing of personal data can be applicable, e.g., the GDPR, the revised BDSG, or the future ePrivacy regulation. Lawful data processing requires compliance with these specific regulations if they are applicable. This makes it difficult to create a “one-size-fits-all approach” for the legal challenges of the platform. These broad changes in content, enforcement, format, and applicable law create severe legal uncertainties [Mü17].

---

3.4 Technological Perspective

A platform realizing our envisioned myneData ecosystem must attract users and data processors and adhere to legal requirements. To this end, the platform must gain users’ trust and answer queries timely and accurately. Furthermore, the platform must provide easily operable user interfaces. Finally, legal requirements demand a secure data management.

Users must trust the platform to protect their personal data from unintended access by third parties and to act according to their sensitivity profile. This also includes that users must be able to actively control how their data is collected and processed. The future legal requirements of privacy by design (cf. Section 3.3) further fortifies this demand. Contrarily, data processors want fast and accurate answers to their queries to gain valuable insights. However, privacy protection mechanisms work by deliberately lowering data accuracy to protect the data of individual users [Sw02, Dw08]. We have to carefully trade these different demands off against each other. The main challenge for the platform's design is thus conflict resolution between requirements from these differing domains.

To attract a large user base, myneData must provide intuitive interfaces. Users must have a clean and simple data control hub and data processors need a good overview of user data shared with them and potential insights. Furthermore, the data collection and querying must be intuitive. While user concerns already strongly motivate a secure data storage, legal obligations may enforce stricter policies than some users would demand. Our platform has to strictly follow these obligations to allow for real-world deployment.

In conclusion, data processors seek fast, accurate insights from large user data pools. However, privacy and legal concerns limit the prospects for data processors. In the following, we present first results for our design towards a sustainable, user-centered data market.

4 Towards a User-centered Data Market

We now give preliminary results working towards a legally compliant data market that is attractive to users and data processors based on the challenges identified in Section 3.

4.1 User Perspective

In accordance to the privacy calculus theory, we assume that users weigh the perceived privacy risks against the benefits for information sharing [DH06]. The evaluation of benefits and risks is influenced by a multitude of factors, such as individual preferences, experiences, knowledge, etc. Other factors are external and depend to a large degree on the exact data-processing situation as they include the types of data involved, trust in the data processor and legal regulations, perceived security of the transaction, and many more. The myneData ecosystem aims to provide control for every user and therefore has to account for these parameters—but no user wants to decide on all of them every time. Thus, the
complexity of privacy control must be scaled down to the most decisive factors while still taking situational and individual preferences into account. To realize this, myneData considers an individual sensitivity profile in which every user can adjust the settings according to her preferences as to (i) which category of information may be shared with (ii) what category of data processors (iii) for what purpose under (iv) what level of privacy protection. A focus group study has shown that these are the crucial factors users want to control [VZL+17]. For users to intuitively adjust the level of “privacy protection,” we are currently developing a privacy control interface in a user-centered approach that is oriented along mental models to be comprehensible for every user.

Trust is crucial and has been studied extensively for other online services. Perceived ease of use, information and design quality, provider familiarity and reputation, third-party guarantees, and many more factors have been identified to generate trust in other online contexts [BDS10, BG08]. The decisive trust factors for myneData and how to appease them still need to be studied further by letting users evaluate a first mock-up of the system.

The other weight on the privacy calculus is the generated benefit. Our user study showed that participants rated the combination of (i) monetary compensation for data provision with (ii) additional information while (iii) staying in control over the data distribution as lucrative. Participants wished to know what data is collected by the services and devices they use, how this data is typically utilized, potential privacy risks as well as the given legal situation. Monetary rewards, e.g., money or vouchers, have been rated as the most wished for type of compensation. However, providing money for data generates defensive reactions as many participants stated, “I do not want to sell my data.” Additionally, some participants stated that the monetary compensation would probably not be high enough.

In a second round of focus groups, participants were introduced to four use cases (c.f. Section 4.2) and evaluated benefits and the intention to use the system. This qualitative study showed people are willing to share data not only for personal but also societal benefits if privacy is guaranteed, e.g., for medical research or donations. Potential benefits for the users are very much tied to the different use cases. In our research, we gained the impression that more tangible benefits within the use cases, e.g., simple access to medical studies, are seen as more profitable than general benefits such as monetary rewards.

We take a user-centered design approach to make myneData comprehensible as well as intuitively and effortlessly usable. Each user is provided with the right amount of information through a multi-layered information design: On a first layer, information is condensed and simple, but the user can access more information on further information layers. Thus, complex privacy statements or consent forms can easily be augmented with graphical or even video-based content as no information is lost. [Sc15]

4.2 Economic Perspective

For the economic perspective, we aim to establish business cases for four exemplary use cases: smart home, medicine, customer insights, and cross service.
Smart Home. We exemplarily consider the processing of thermostat data. Smart thermostats deduce individual parameters of buildings to control the flat heating and save money by reacting accordingly. These parameters may reveal energy-saving potentials that could be achieved if the building was modernized. This valuable insight could subsequently be leveraged by vendors, service providers, or constructors. Finally, heating system manufacturers can use this data to improve their systems as well. Another scenario involves planning office rooms. Planners currently have little chance to monitor building utilization and how previously simulated building characteristics work out in the real world. Door and window sensors, temperature and humidity sensors, and presence sensors can generate valuable insights about usage patterns and help improve future buildings.

Medicine. In the medicine sector, myneData can aid medical surveys by offering an agent between pharmaceutical companies and study participants. Those two customer groups are matched via the myneData platform and during the study participants can track their study-related information about daily chores. This ensures a genuine overview of the collected data for the companies and eases the process for both customer groups.

Customer Insights. The main purpose in this case is the customer’s control of data transfers. The customer can start collecting data within myneData, for example, by linking social media websites or by installing web trackers. Then, the customer can decide to sell selected pieces of data to interested companies. The companies consider those data pieces as highly attractive as they can gain insights about different customer segments and improve their marketing strategies. Meanwhile, the customer is able to get an overview of his or her data and see his digital “footprint” that is established on the Internet.

Cross Service. The use case “Cross Services” addresses service companies or, more precisely, call centers who offer call center services for a variety of firms. Call centers are neither allowed to save personal information about customers nor to recommend products of other clients to their customers (a cross service). With myneData, we provide an option for call centers to bypass this restriction. The customer can establish a myneData profile and allow the call center to save information on him/her in that profile. Each time the customer calls the call center, they can choose to grant access to their profile and the call center can then provide much better consulting, even w.r.t. to other products if the customer desires so. The benefits for the customer are not only the improved consulting but also the power over their own data, since they are the manager of their data cockpit.

4.3 Legal Perspective

Legally, data processing as performed in myneData falls under the term “profiling.” According to the definition in Art. 4 No. 4 GDPR, this is any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyze or predict aspects concerning that natural person’s performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location, or movements. Although Art. 22 GDPR addresses the process of profiling, it does not provide a legal basis for its conduct. Rather, the norm
sets up special requirements for controllers and specific rights for data subjects when profiling is performed [AJ16]. Therefore, the data processing in myneData has to be based on a general legal permission, i.e., consent or a legal provision in the interest of the data processor. However, which specific permission applies to a processing scenario depends on the source and category of data processed as well as the situation of data collection. In the different scenarios of myneData, these circumstances may vary. As any handling of personal data is legal if the data subject agrees to it (Art. 6 (1) lit. a) GDPR, Artt. 6 et seqq. ePrivacy Regulation Proposal), data processing in myneData will be based on users’ consent. This way, all processing in the project refers to a coherent basis. Furthermore, given user consent may even allow to collect, analyze and process special categories of personal data, i.e., sensitive data such as information about a person’s health, sexual orientation, political opinion, or religious beliefs (Art. 9 sec. 1 GDPR). Considering these advantages and the current legal uncertainties concerning the interpretation of the GDPR, consent as permission for data processing will generally become important during the transition period and until a deeper understanding of the GDPR has been gained.

Rights of data subjects in the GDPR require the technical possibility to access, correct, erase, restrict, and port personal data at any time (Arts. 15–20 GDPR). Moreover, already the design of a platform has to ensure compliance to controllers’ duties concerning the execution and configuration of the data processing, such as principles relating to processing of personal data (Art. 5 GDPR), duties to inform (Arts. 12–14 GDPR), or security of processing (Art. 32 GDPR). Consideration of these requirements at such an early stage can ensure that the platform provides the technical conditions for their fulfillment. This approach has gained special importance under the new European framework. Art. 25 sec. 1 GDPR obliges the controller to implement appropriate technical and organizational measures designed to implement data-protection principles and to integrate the necessary safeguards into the processing in order to meet the requirements of the GDPR and protect the rights of data subjects. Under Art. 25 sec. 2 GDPR, the same applies to technical and organizational measures to ensure that—by default and thus by technical measures—only personal data necessary for the specific purpose is processed. Infringement of these principles “privacy by design and by default” may lead to fines.
4.4 Technological Perspective

Users and data processors come together at our envisioned data market to trade personal user data. Figure 2 depicts our proposed architecture for the myneData platform. Its main component is the data cockpit, where users manage their data sources and sensitivity profile and where they review and approve data processing requests. Furthermore, data processors have a high-level overview of their pending and past data processing requests. Note that the platform is only logically centralized. For especially critical data, e.g., medical data, we currently investigate decentralized alternatives to our design as well.

The myneData platform performs data collection and processing in six steps. First, a user authorizes the platform, and thus gives consent, to collect the data from her data sources. Here, it is challenging to integrate various data sources with diverse APIs. New data sources can gradually be integrated due to an extensible data API that transforms data into a general-purpose data format. Subsequently, data processors query the data pool via an intuitive query API. They can request anonymized datasets or statistical information such as the average activity level among local students. Following the request is the query’s approval phase: Users who can provide valuable data for the query are asked to share this data with the data processor. Each user can review the terms of the query, i.e., which data is requested by whom, the number of expected participants, how it will be processed, and how participating users are compensated. The user can thus willingly accept or deny this request. A dedicated authentication module with a payment submodule monitors that no data leaves the platform until the query’s terms are satisfied. Once enough users decided to participate, the platform computes the result by processing the data and applying privacy-enhancing technologies such as $k$-anonymity [Sw02] or differential privacy [Dw08] to protect the users. Finally, the platform responds to the data processor by notifying her and returning the obtained result in a standardized format such as a JSON-based object.

One main challenge for our platform will be to resolve conflicts between user and data processor interests. As a sufficiently sized user base is critical to an online platform’s success [Mu15, HH14], we prioritize user concerns over data processor interests whenever conflicts arise. We believe that this high-level concept fits the needs of both users and data processors when it comes to realizing a transparent, easy-to-use data market.

5 Related Work

Previous work from different strings of research tackled aspects of myneData. Commercial platforms such as Powr of You (powrofyou.com) or Datacoup (datacoup.com) incentivize users to sell their data. Recently, we notice a shift towards platforms that give users control over who can access their data [KPH15, HHM+14]. These existing platforms provide up to two out of consent management, self-monitoring, monetary rewards, and privacy w.r.t. third parties. With myneData, we seek to combine all four properties within one platform. To achieve this, we must thoroughly investigate user demands. Users’ trust,
privacy attitudes and behaviors have also been extensively studied in different online contexts in the last decades, oftentimes reaching the same conclusion: User decisions are individual, partly irrational, and largely dependent on the context [KWM13, ABL15, SDX11]. Therefore, empirical studies investigating trust factors, privacy preferences, and desired compensations specifically for the case of the myneData ecosystem are essential.

Moreover, there have been different approaches to examine the legal foundations of commercial profiling and, especially, to find a suitable legal definition fitting the process [Sp16]. These solutions, however, have been proposed for the specific purposes of highly specialized projects so it is debatable that they can be applied on a more general basis.

6 Conclusion

We introduced our vision of myneData, a sustainable and user-controlled data market. Our goal is to enable users to share data with third parties without privacy loss. We identify challenges for our platform that arise from user and data processor demands, the need for a business model, and future legal requirements, e.g., the new European General Data Protection Regulation. Our initial results outline a data-sharing ecosystem that requires a comprehensive understanding to realize myneData: First, we base our user model on studies affirming the privacy calculus theory. Second, we identified four use cases that allow SMEs to benefit from myneData and strive to derive a business model from that in the future. Third, we identified upcoming European law on data privacy as both a challenge and an opportunity for the introduction of a privacy-aware data market. Finally, we outline an architecture for a trusted myneData platform. In the future, we plan to investigate distributed alternatives to this platform in order to further decrease the trust required by users.

Acknowledgements

This work has been funded by the German Federal Ministry of Education and Research (BMBF) under funding reference numbers 16KIS0443 to 16KIS0447. The responsibility for the content of this publication lies with the authors.

References


Roman Matzutt et al.


