MASTER'S THESIS

The Future of Personally Identifying Information Ownership

Ewijk van, R.

Award date:
2021

The Future of Personally Identifying Information Ownership

Degree programme: Open University of the Netherlands, Faculty of Management, Science & Technology Business Process Management & IT master's programme

Course:	IM0602 BPMIT Graduation Assignment Preparation IM9806 Business Process Management and IT Graduation Assignment
Student:	Rene van Ewijk
Identification number:	
Date:	29 October 2021
Thesis supervisor	Laury Bollen
Second reader	Rachelle
Version number:	1.0
Status:	Final

Abstract

The General Data Protection Regulation was adopted in Europe in 2016 and instituted in 2018. With the adoption of the General Data Protection Regulation, European citizens have gained more control over their personally identifiable information in an attempt to interrupt the monopoly of the “big tech” companies. Currently, European citizens do not share in the value obtained from the use of their data. Accordingly, this study focuses on the possibility of introducing a single digital identity to aggregate all digital personally identifiable information by inviting content experts to participate in a Delphi study regarding the future development of the ownership of that information. This thesis examines the potential future of managing personally identifiable information by following current trends in four key areas: compliance, security, trust, and privacy.

Compliance: Moving from the current administrative framework to a technically centralized managed technology will simplify compliance and decrease the administrative burden companies currently experience.
Security: There are security concerns in the introduction of single-points-of-failure when introducing middleware. Fortunately, mitigating technologies have already been developed to negate this risk, which are a mix of encryption hashing and peer-to-peer transactional technologies such as the IMRA, Chainlink 2.0, or Sovrin networks.
Trust: Trust can be established by combining decentralized blockchain-based technologies and centralized service providers such as governmental services. The technology around ensuring privacy and trust is already mature enough for adoption.
Privacy: Any future for centrally managing personally identifiable information must include a privacy by design approach. Technologies such as zero knowledge proof appear promising with regard to protecting the data subject’s privacy while interacting with service providers.

The trend in Personal identifiable information ownership is moving more towards the data subject’s ownership. This concept is gaining momentum since European legislation is pushing for data ownership to make the data available to European companies in order to create more competitive European companies. Combining decentralized technologies together with centralized trusted identity authorities introduces new business models in managing Personal identifiable information similar to developments that can be currently seen in the banking sector.

Keywords: Future of personally identifiable information (PII), General Data Protection Regulation (GDPR), identity management, hybrid identity model, blockchain, DNS-IDM, Chainlink 2.0, Oracle network, Delphi study, data ownership, big tech monopoly, decentralized technologies, centralized technologies, trust, privacy/security, GDPR compliance.

Table of Contents

1. Introduction

1.1 Background

On the 12th of March, 2014, the European Parliament adopted the General Data Protection Regulation (GDPR) to update the 1995 Directive 95/46/EC. The GDPR introduces new rights for data subjects, which are intended to rebalance the power distribution between the citizens of the EU and large global companies (Auwermeulen, 2017).
Powerful international organizations offer a number of online services that provide consumers with considerable advantages. These organizations offer innovation and efficiency in problem solving and are accessible to many people in Europe (Supervisor, 2014). However, these companies also use the consumer data they collect through the services they offer for commercial ends. The European data protection supervisor has stated the following:

> The rapidly expanding online market or markets…increasingly touch all aspects of business. Making sure competition works effectively in these markets will be a major priority…the growing collection, processing and use of consumer transaction data for commercial ends…is proving an increasingly important source of competitive advantage [which could be] an increasing source of consumer detriment.
> (Supervisor, 2014)

Access to the online services that are provided by international companies (service providers) often require the disclosure of personal information. The acquisition of this information enables international companies to better answer customer needs and deliver better quality. Hence, this personal data is extremely valuable (Damien Geradin, 12 February 2013).

Consequently, personal data has become a key resource for online service providers. Concerns have been raised about the acquisition and processing of data by service providers who raise the barrier of entry and purposefully monopolize the market (Damien Geradin, 12 February 2013). Because customer data is at the core of their business model, service providers are trying to retain their grip on this data by making it purposefully burdensome or expensive for users to shift to a new service (Auwermeulen, 2017). This behavior results in a lock-in effect that creates a high risk of market abuse (Engels, 11 June 2016).

The use of big data and big data analysis encompass at least four dimensions (Customer engagement, cost reduction, decision making and asset optimization) that create a competitive advantage for those who have access to that data (Shan, Luo, Zhou, & Wei, 2019). It is easy to imagine how service providers can leverage these dimensions to their advantage, which also means that organizations who do not have access to this data do not have this competitive advantage.

Nonetheless, the core of the GDPR is aimed at privacy:

> The protection of natural persons in relation to the processing of personal data is a fundamental right. Article 8(1) of the Charter of Fundamental Rights of the European Union (the 'Charter') and Article 16(1) of the Treaty on the Functioning of the European Union (TFEU) provide that everyone has the right to the protection of personal data concerning him or her. (UNION, 2016)

Article 20 of the GDPR, the right to data portability, was introduced to mitigate the concern that the processing of subjects data leads to the previously mentioned lock-in effect (Engels, 11 June 2016). The article leaves data subjects in control of their immaterial wealth and the privacy from which service providers currently profit (Paul De Hert a, 2018). Data portability warrants the data subjects control of their data when processed by controllers. Hence, data portability is located on the intersection between data and privacy protection – the right to own and control one's own data – and other fields of law. The GDPR positions the data subject as the controller of their data and constitutes a case for privacy enhancing technologies that allow individuals to enjoy the immaterial wealth of their personal data within the data economy (Paul De Hert a, 2018). The question then arises of whether this could be achieved by introducing a middleware solution to manage a person's information and privacy.

1.2 Exploration of the topic

The GDPR and big tech market dominance

The scope of the **GDPR** spans any organization that collects or processes information related to European citizens, no matter where the organization or data is based. It introduces the obligation to notify the authorities of a data breach within 72 hours and extends personally identifiable information (PII) directly and indirectly to include identifiers such as IP addresses and cookies (a small text file stored on your computers browser program for server side browsing functionality) (Tankard, July 2016).

The European Commission is concerned about the current monopoly of big tech with regard to PII and would like to change the current power dynamic by introducing new legislation to improve the level of competition. This concern demonstrates that there is a secondary motivation for implementing the GDPR (See Appendix 1.1 GDPR European motives).

Personally identifiable information refers to

> any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.(Supervisor, 2014)

This kind of data has become a resource for online service providers (Damien Geradin, 12 February 2013). **Big data** makes information more transparent. It helps to make data more transactional in a digital form and create more precisely tailored products, enabling the producer to appeal to more people. The data is used to make management decisions. It also enables the next generation of services and products, which means that access to this data is imperative for innovation (Tim McGuire, 2012).

The control of data impacts the power structure and business models of the existing tech companies and provides those companies who own the most data with a competitive advantage. Changes in control of that data should lead to new opportunities for service providers and create competition (See Appendix 1.3 Big data control).

Competition has long been of central importance to the European Commission. Competition law concerns the behavior of companies and the abuse of market power. This concern has evolved from focusing on preventing public obstacles to interstate trade but now seeks to ensure there are controls on corporations with the aim of making the market more efficient and available to its citizens (Damien Geradin, 12 February 2013). Big tech companies (such as Google) have been shown to have abused their **market dominance** under European law. These **monopolies** have created an unfair competitive advantage within the European market (Gergely Alpár & Bart Jacobs, 2017).

Consequently, some of the articles within the GDPR focus not only on privacy but also on competition, creating new opportunities for service providers and transferring legislative power over data away from the Big Five tech companies (Google, Facebook, Amazon, Microsoft, Apple) and returning it to European citizens. The transference of data ownership to European citizens is likely to lead to **new business models**, thus, creating more **competition** (See Appendix 1.2 European competitive advantage and Appendix 1.4 Business models).

Middleware and PII management

Many potential technologies are available to support the change of data ownership from the service provider to the data subject (See Appendix 1.5 Technologies). Some of the middleware solutions (software that facilitates communications between two or more applications or components on a distributed network) that have been developed are DigiD and “I Reveal My Attributes” **(**IRMA).

DigiD is part of the Dutch E-government strategy that begins with the Basis Registratie Personen (BRP). This BRP provides the government with a citizen's "source identity," which is the basis for any other identities with a bank, telecom company, and so on. DigiD is an authentication layer that Dutch citizens can use for authentication with municipalities and other governmental services. These services can then request the data via DigiD that the service requires to operate (e.g., tax returns, moving house). However, DigiD is limited to the government domain. DigiD as a central hub is considered a trusted source for identity management because of the integration with other governmental systems such as the BRP. Metadata at DigiD is used primarily for anti-fraud monitoring and is not traded commercially (Jacobs, 2015).

eID: Elektronische identiteit is a new attempt by the Dutch government to create a new authentication system that could connect governmental and commercial services. There are still significant concerns regarding privacy and security. Most of these concerns relate to the traceability of the authentication, using PII for authentication provides services with the power to track every transaction, and the security of the application. The more data there is to gain, the bigger the target for malicious actors(Jacobs, 2015).

The Privacy by Design Foundation created **IRMA** in 2016. IRMA is an application that applies attribute-based credential (ABC) principles in practice to protect the user's privacy and return control over PII to the data subject. IRMA is an identity management tool that collects a user's PII and make its available to service providers when necessary (Foundation, 2019). For example, when an individual purchases a bottle of whiskey online, the retailer only needs to know whether the purchaser is 18 years or older. The retailer does not need to know the purchaser's gender. With IRMA, only the information required by the service provider can be shared. Full control of the data remains with the data subject. See Figure 1. IMRA architecture

An ABC ecosystem contains the credentials that can be used by many service providers. This is accomplished by using the IRMA API and service provider API to exchange information. The credentials are managed within a wallet-like smartphone application, which gives the user centralized control over their information (Gergely Alpár & Bart Jacobs, 2017). However, because the exchange of information is not natively supported by IMRA and service providers, only a few services and integrations are currently available on this platform (Foundation, 2019).

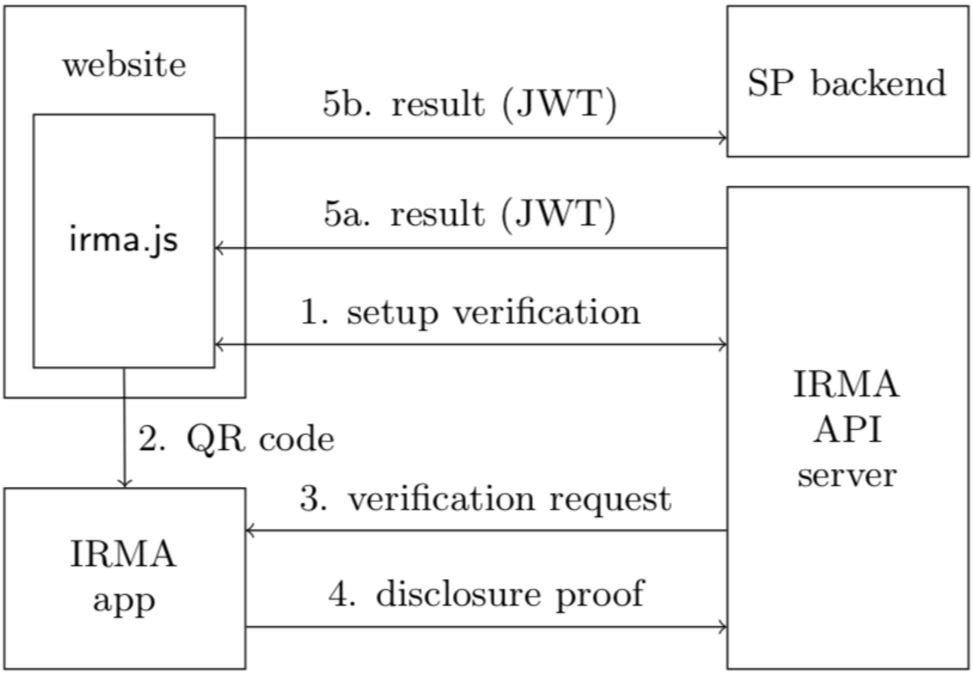

Figure 1. IMRA Architecture (Foundation, 2019)

1.3 Problem statement

With the introduction of the GDPR into Europe, data controllers and data subjects are attempting to find their way in this new legal framework. The GDPR is designed to address many of the EU‘s concerns regarding data privacy, which began in the early 1990s. In his research, Paul de Hert states that “This represents the first theoretical step towards a default ownership of personal data to data subjects”(Paul De Hert a, 2018). This ownership has now been implemented into law.

However, three problems have been identified with the GDPR: First, the ability of service providers to implement the GDPR and, second, the ability of consumers to benefit from it (1.3.1). An extension of the latter point is that EU consumers do not share in any of the value obtained from the use of their data (1.3.2). Third, consumers do not control their own PII, and there is a lack of direction when service providers implement the GDPR (1.3.3). These problems are detailed in the following sections.

1.3.1 Data controllers struggle with GDPR implementation, and the rights of the data subjects are unclear.

Abuse of PII: Personal data is used as a resource for online service providers. Service providers have demonstrated their misuse of this information multiple times (Damien Geradin, 12 February 2013).

Enforcing the GDPR: Confusion exists regarding the various rights of data controllers under the GDPR (Wong, 8 October 2018).

Lack of security and responsibility in PII processing: Data is processed not only by the service provider but also by sub-processors, which means the PII is vulnerable to misuse and out of the data subjects’ control. It is often unclear – even to the service providers – where and how data is processed (Cellerini & Lang, 2018).

Data breaches: Data breaches have become commonplace in Europe to the extent that many legislative actions have been taken, including fines to encourage companies to secure the information of EU citizens (Cellerini & Lang, 2018).

1.3.2 EU citizens do not share in the value gained from their data

Lack of competition: There is not enough competition in the European digital market. It lacks innovation and tech companies are unable to develop more competitive and innovative products. Due to the market dominance by the big Five tech companies. This is of great concern to the EU (Paul De Hert a, 2018). This leaves the consumer in Europe with less optimal digital services.

Service lock ins: Currently data subjects are still mostly locked in to the service they use. The right to data portability was introduced to mitigate the concern that the processing of a subject’s data leads to a lock-in effect (Engels, 11 June 2016). However, the effect of Article 20 has yet to be studied. Current business models and service providers do not seem to show it has had an impact on locking-in data subjects. This situation removes the flexibility of consumers when seeking other services.

Citizens do not profit from their PII: The main driver of value for the big tech companies is the ownership of large quantities of data on consumers. Citizens are not in control of the data from which service providers currently profit (Paul De Hert a, 2018).

1.3.3 Lack of direction in middleware identity management

Currently deployed identity systems, that is, systems that house PII, have a **dictatorial approach in terms of managing PII**. The service providers in charge of maintaining and controlling personal data do not always follow the GDPR guidelines (Damien Geradin, 12 February 2013). In addition, there is **a lack of trust** between the identity provider and the data subject, as demonstrated in 1.3.2.

Identities are not managed and owned by their rightful owners: Service providers are managing PII rather than the data subjects, and it is a challenge to securely control and manage PII (Jamila Alsayed Kassem, 2019).

1.4 Research objective and questions

Privacy has become a frontline issue in Europe. With increasing awareness of privacy issues, new questions are also arising concerning competition and how data subjects interact with digital services. A first step has been taken by implementing the GDPR.

This research focuses on a middleware solution for PII management. It does so by focusing on new ways consumers interact with technology while introducing a new GDPR concept for PII ownership. The research objective is to add a substantive theory to the body of knowledge by researching three aspects of middleware that could be used to manage PII (ownership) in the future. These three aspects of investigation give rise to the following research questions:

RQ1: *How can the introduction of middleware help service providers with GDPR compliance?*

RQ2: *What are the main factors in establishing trust to facilitate the complete ownership of PII using middleware?*

RQ3: *What are the security concerns involved in introducing a middleware solution to manage PII and how can these be addressed?*

RQ4: *What are the main factors in the privacy domain when introducing middleware to aggregate all PII?*

1.5 Motivation/relevance

The motivation for this research is to analyze the trends and examine one aspect of the trend in privacy legislation, namely the effort of the European Commission to improve competition and innovation in the European digital market. The European Parliament seeks to increase market competition in the data economy to enable the development of more user-centric platforms for the management of personal data (Paul De Hert a, 2018).

The possibility of consumers controlling their own data using a middleware solution, and profiting from that system, could change the internet as we know it. This fact alone is exciting and worth investigating. Many different business model opportunities for changing markets exist, as has occurred in the financial service industry by opening up financial data to third parties (Vanberg, 2016).

The innovation that might develop from changing the current business models could lead to the diversification of consumer choices and new innovative services. Consequently, this change presents a stimulating opportunity to research into identity management. The idea of users controlling their own data might return power to consumers and change the way they interact with digital services.

The purpose of this research is not to build a system, only to focus on three aspects of middleware systems and to examine how privacy, security, and trust will impact the introduction and use of a middleware solution for one single digital identity management system. These aspects are intended to address governmental concerns regarding service providers and help in the development of a middleware solution to facilitate the future enhancement of PII. Ultimately, the aim is to establish exciting new consumer services while consumers enjoy the ease of having a digital identity.

In this regard, developments around the COVID-19 pandemic have led to a situation in which vaccination passports are being discussed. Christopher Dye noted, “There are many issues surrounding the fair use of vaccination passports” (2021). The concept of the COVID-19 passport is an ideal example for this research because it involves sensitive PII that would need to be shared with service providers. For this reason, this case is briefly discussed in relation to how PII management of the data might be undertaken.

For such a passport, a device to support data would be required, which could be a chip or a smartphone. When someone is vaccinated, they would register their claim at a trusted authority, such as the national health service. This claim would then be transferred to the person’s digital identity – hosted on thousands of different nodes owned by different authorities – and hashed to ensure that unauthorized access to the claim is not possible.

Following this, for example, the person in question may then wish to travel and subsequently requests to book a flight. The service provider would like to ensure that this person adheres to certain policies, such as being above 18 years, having a minimum of 300 euros in the bank, and being **vaccinated.** This person could then use their digital identity to verify those claims. In this case, the only data that the user needs to exchange is **whether** they adhere to the conditions named above. They do not need to supply their age or the amount of money they have in the bank. The service provider accepts the claims (since they are verified by a trusted source) and provides the service that this person has requested.

Chapter one introduces the topic. Chapter two discusses the relevant literature. In chapter three the Delphi study is designed, and the results are discussed in chapter four. In chapter five the conclusion of this study can be found including the limitations of the study.

2. Theoretical framework

2.1 Research approach

The purpose of this research is to understand the trend of PII ownership in Europe by using a middleware approach and to ascertain what the opportunities and barriers are to transferring data ownership from the service provider to the data subject to place the data subject in a position of power.

A digital identity is used to authorize access to information systems and ensures that only valid users gain access to the right systems. This identity has become more significant for the development of services and technologies that consume PII (Domingo, Madrid, Spain, 2018.). *Can **centralized** or **decentralized middleware** facilitate the PII exchange between data subjects and service providers?* The advantages and disadvantages are discussed in 2.3 Results.

The GDPR addresses service providers privacy and security concerns and "the protection of the natural person in relation to the processing of personal data is a fundamental right" (Supervisor, 2014). *Are there any **security** and **privacy concerns** regarding the use of a **middleware** solution to manage PII between data subjects and service providers?* The privacy and security aspects are researched in the context of the GDPR in 2.3 Results.

Internet services demand a scalable **solution** for **trustworthy identification**. There is currently a shortage of proficient **trust** management schemes for online services in large-scale adoption computing paradigms by the public (Mukalel Bhaskaran Smithamor, 21 April 2018). Ways in which trust can be established when PII is exchanged are discussed in 2.3 Results. *"Because of increased technological complexities and multiple data-exploiting business practices, it is hard for consumers to gain control over their own personal data" (Vrabec, 11 December 2018).* Due to its complexity, it can be difficult to implement the GDPR. Can ***middleware*** *help with* ***GDPR implementation****?*

Background and problem statement

Mixed research techniques are used to develop the research questions and problem statement. Different angles of the GDPR have been studied, mainly focusing on the connection between GDPR Article 20 and privacy regulation, and the EU's desire for more competitive companies within EU boundaries. The current problems with the GDPR and the connection between the European Commission and Article 20 were researched using database searches regarding the Big Five, how they behave and what their relation is to the EU. Qualitative data was used to develop quantifiable problem statements. Qualitative data was used to establish causality between the problem statements and their subjects regarding the GDPR.

The review began with the GDPR articles by Barbara Van der Auwermeulen, who examined the notion of data portability and possible motives for its introduction besides data security. Many other authors were found who were critical of the motives of the EU Commission's implementation of the GDPR. Expanding on the work of those authors, a list of keywords [Future of personally identifiable information (PII), General Data Protection Regulation (GDPR), identity management, hybrid identity model, blockchain, DNS-IDM, Chainlink 2.0, Oracle network, Delphi study, data ownership, big tech monopoly, decentralized technologies, centralized technologies, trust, privacy/security, GDPR compliance was then used to develop a search strategy to identify the relationships between the topics. The selected key words and phrases were identified in multiple databases and used to expand on each topic when necessary.

Conceptual Framework

To obtain the literature required for the theoretical framework, a list of studies related to the research subject was created. The quality of 90 papers was assessed using diagnostic and summative techniques. Relations and contradictions were explored and documented. After the list was distilled to 10 papers, these papers were used to build up the arguments that comprise the theoretical framework. The work of Jamila Alsayed Kassem was particularly influential and was used as a means to retrieve other relevant articles and studies. Databases such as EBSCO and Google Scholar were used for this research.

More unconventional sources were also explored to create more context for the theoretical framework. Podcasts about the subject were summarized, some experts were spoken to, and several professional interviewes (Arie Juels, Herbert Blankenstijn, Lex Friedman and Jaron Lanier) were used to gain a better understanding of the field. Furthermore, a forum about identity management was visited where insights were gained by networking and talking to leading industry professionals. This information was used to create a conceptual model of the topic. The search for articles began based on the conceptual model Appendix 1.6 Conceptual Model. Figure 1

At the core of the model was the GDPR. Many articles from the European Commission website were retrieved to lay the foundation for what the European Commission has attempted to achieve with the regulation.

In terms of technology, the IMRA Foundation has a section on scientific literature because one of the objectives of the foundation is to promote scientific research on this topic. Those articles were used to form the basis of the theoretical framework.

2.2 Literature Review

Middleware to facilitate PII exchange

When exchanging data on the internet it is critical to be able to establish the user's digital identity to authorize access or process certain information. Currently most of those exchanges are carried out using an identity management system that employs customer identity and access management (CIAM) systems. Identity management enables the customer to take control of their data using third party tools (Rasouli, 2019) and is a method of validating and recognizing the user. Once validated, the user is granted access to confidential information.

The identity provider (IDP) manages the information, while the relying party (RP) delegates the responsibility of authorizing a user to the IDP. The user experience is leveraged by offering the user a single identity that is valid for different services (Fett, 2017). The **centralized schema** offers a single and centralized system for the user, which is achieved by the RP establishing trust with the IDP. The RP accepts the user's claims (about the user's identity) from the IDP and authorizes the user on the platform (Fett, 2017).

Centralized systems

Service providers such as Google and Facebook have implemented the "social" management of identities and, in turn, have become centralized identity providers for a large number of users by offering a single sign-on. These services increase simplicity and ease of access for the user by implementing CIAM principles (Rasouli, 2019). These are also the service providers who are currently leveraging PII for a competitive advantage (Appendix 1.3).

The single sign-on is accomplished by the IDP using Security Assertion Markup Language (SAML) and Open ID 2.0 (Appendix 1.5 Table 1). However, these services exchange credentials with different layers of communication, making them vulnerable to malicious identity provider attacks, malicious RP or Service Provider, and replay attacks. Exchaging credentials with different layers of communication degrades security and privacy standards by leaving PII vulnerable on the internet (Jamila Alsayed Kassem, 2019).

The Dutch DigiD is an example of a centralized system. With DigiD, people can easily, securely, and reliably authenticate their identities for digital services. The user is presented with a login screen with the choice of authentication method. DigiD operates as an authentication service using means of digital authentication and offers extensive support for users and services (Logius, 2016). The most obvious difference between DigiD and a CIAM system from, for example, Google is that the DigiD identity is linked to the BRP, identifying the user as a natural person. A process in place whereby the user needs to request access and be verified with a passport. Web services can be integrated with the DigiD system. In general, these are health- and municipality-related services.

The user interacts with DigiD – the middleware core – that delivers secure authentication and authorization for the user. After the verification of the user, the DigiD security token can be used for authentication with integrated webservices (Logius, 2016). The logical DigiD dataflow is represented in Figure 2. DigiD Dataflow

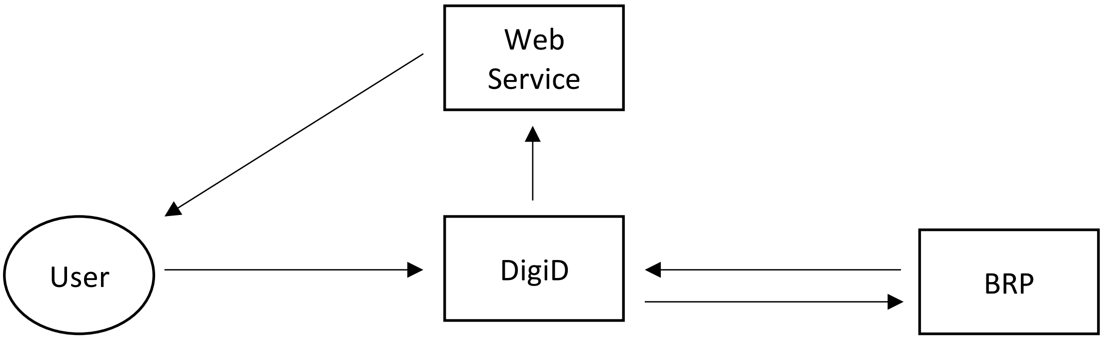

Figure 2. DigiD dataflow

There are, however, concerns that the currently deployed **centralized systems** maintain third party control of the data and PII. Consequently, the challenge is to securely manage and protect PII that is managed by central institutions and humans who are responsible for the entire activity. Accordingly, the digital identity is not controlled by the data subject but by the provider. See Figure 3. Centralized identity management

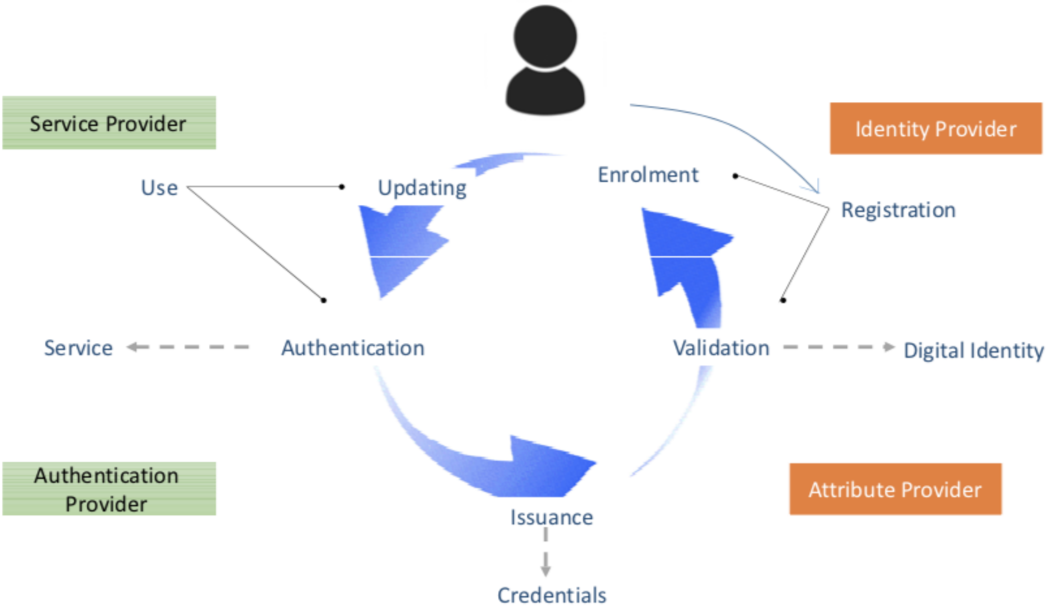

Figure 3 Centralized identity management (Fett, 2017).

Decentralized systems

Decentralized systems have gained more traction since the launch of Bitcoin in 2009. The underlying technology that is used for Bitcoin is **blockchain** technology. The blockchain ledger enables developers to deploy decentralized applications (Dapps) that are not bound to any centralized control (Treasury, 2019).

Ethereum is a protocol built on blockchain technology that provides a tightly integrated end-to-end system for implementing applications. Ethereum is based on the use of **smart contracts** and is available as an open-source protocol on which to build identity systems. Ethereum has a robust infrastructure and natural security features such as distributed denial-of-service prevention, and a serverless and decentralized infrastructure (Triantafyllidis, 2016). Dominant identity management systems in this space are Serto, Sovrin, ShoCard, Cambridge Namecoin, Blockverify, and Cambridge Blockchain.

The Sovrin Network ingests claims about a user from a certified agency such as a bank, a school, or a municipality. Those institutions have inherent user trust because of their stature. These institutions connect to the Sovrin network. On this network, the user can login and assemble their claims from different institutions. The user can then choose to use any of those claims to verify themselves for other services; for example, using one’s bank account number to facilitate a transaction with a third party vendor(Sovrin, 2020). A network example can be seen in Figure 4. Sovrin network.

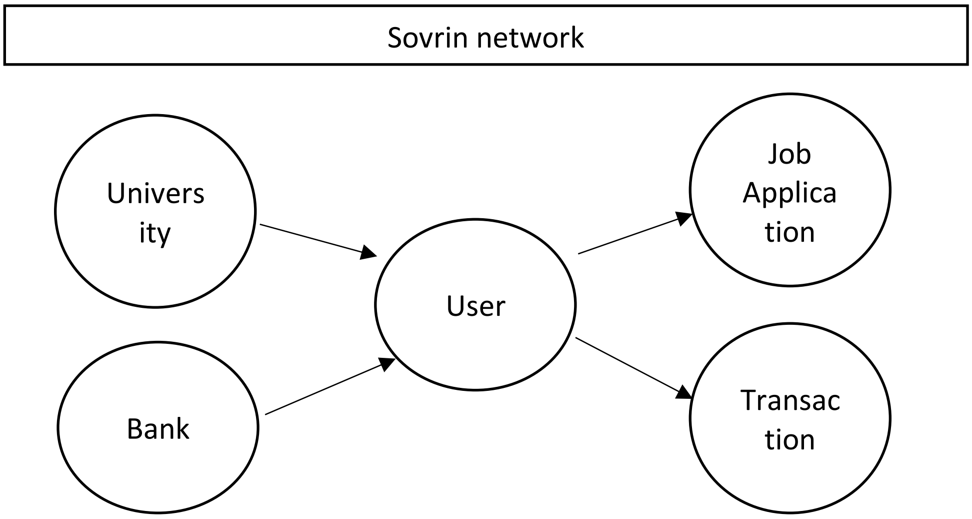

Figure 4. Sovrin network

The domain name service-identity management system (DNS-IDM) is an identity system based on a smart contract identity system. This system administers identities and associates PII with certain attributes, returning governance to the data subject. In current research, DNS-IDM systems outperform centralized systems in terms of privacy and security due to their decentralized natures and are able to address the current limitations and threats that centralized systems experience (Jamila Alsayed Kassem, 2019).

This system offers a domain name system-like experience for the management of PII. A data subject can make use of the real-world attributes of their identity, which can be seamlessly utilized by service providers to authenticate and authorize the user for certain privileges within the service provider's system. This exchange is completely facilitated by blockchain technology, which means that the information the service providers consume only remains valid as long as the data subject wishes (Jamila Alsayed Kassem, 2019). In current centralized systems, third party claims about an identity attribute remain valid once they are verified. Blockchain-based identity services rely on the claims the data subject discloses based on their acceptability and enable the data subject to maintain direct control over their PII.

Privacy and Security

Using middleware for GDPR compliance and to protect a natural person's identifiable information does introduce some privacy and security concerns. These concerns depend on the nature of the system. Centralized identity management middleware introduces different concerns to those of decentralized identity management systems.

When discussing **centralized** identity management systems, the most prominent CIAM systems are the previously mentioned social identity systems implemented by Twitter, Google, and Facebook. The main issue for these systems is that centralized identity management systems deprive the user of the actual ownership of their online identity. The IDM needs to adhere to GDPR principles, but the user has no real control over their PII and has to trust the privacy and security governance of the centralized system. Subsequently, with this obvious lack of user control, the centralized system can easily abuse the PII by granting access to third parties to the PII. These services also tend to consume more PII than is needed and permission to do so is buried in the terms and conditions that most users do not read. This level of consumption raises privacy and security concerns because the nature of centralized systems means they are vulnerable as a single-point-of-failure for honeypot DDoS or other malicious attacks (Jamila Alsayed Kassem, 2019). Centralized systems use JavaScript redirection, which is used to exchange the user's claims. The layers of communication that take place for authentication make the communication vulnerable to identity provider, service provider, and replay attacks (Jamila Alsayed Kassem, 2019). There is also a lack of incorporation and integration between different IDMs from different vendors, making it more challenging for a data subject to gain control over their PII (Ferdous, 2012).

Initiatives have been introduced to remedy the flaws of the current centralized model. The ABC principles developed by IRMA only reveal the particular attribute a service provider requires. This technology is also based on blockchain technology and returns control over PII to the data subject (Foundation, 2019). This initiative allows systems to prioritize privacy and restrict access to further information by service providers. This technology uses IRMA as an identity provider thus introducing a privacy friendly variation of a centralized identity management system Figure 1. IMRA Architecture.

The **decentralized** model built on blockchain technology gives the user complete control over their identity. All privacy concerns are addressed because the user can maintain ownership of their identity on a secure ledger. Because the identity is housed in the ledger, it is not fragmented between multiple systems, thus, removing the single-point-of-failure security-related concerns. This model also removes the concern of data being vulnerable through data breaches at service providers.

The decentralized model also addresses phishing concerns and removes the need to trust a centralized system to secure one's data and service providers who bury outrageous provisions in their terms and conditions regarding access to users' PII (Jamila Alsayed Kassem, 2019). Nonetheless, a decentralized system also has limitations. Because the blockchain technology is based on encryption, advances in hardware may render the security of the blockchain vulnerable to breaches. The lack of validation on the blockchain (because of the exchange of public key/private key) may expose more data then intended (Jamila Alsayed Kassem, 2019). Depending on how the DNS-IDM system is built, different privacy and security problems may be exposed.

Trust in a digital transactional environment

Bhaskaran states that "The establishment of trust associated with any cloud service depends on the performance evaluation of that service. In-depth and accurate assessment of cloud service performance is necessary for both consumers and providers" (Mukalel Bhaskaran Smithamor, 21 April 2018). A fundamental level of trust is required for digital entities on a network to perform a transaction (Jamila Alsayed Kassem, 2019). Because the digital identity becomes a proxy for real-life, trust in the identity is essential.

Trust in a centralized environment lies completely with the service provider. The service provider relays identity trust to other service providers, which represents a back-end system where the identity is relayed between systems (Jamila Alsayed Kassem, 2019). One recurring theme is that implicit trust is given by the user when relaying their identity to the service provider. Trust in a centralized system is two-fold:

1. Trust in the central system relaying PII from the user to the service provider
2. The trust the user has in the IDM (by giving their information to the IDM)

Centralized systems such as Facebook and Google have low levels of public trust but billons of users.

Trust can also be established using a governmental authority. In the Netherlands, the Dutch government administers PII for its citizens digitally and provides them with access to a variety of services (Jacobs, 2015). This trust could work in a central model.

Trust can also be found within decentralized models. When the PII is managed in a decentralized way, trust is implicit because the data subject is in complete control of the ledger containing the PII. A decentralized network removes the necessity for trust in a centralized authority. Trust in the technology supports the decentralized system because it needs users to function and the self-ownership and control over the PII that the service providers consume. The whitepaper Chainlink 2.0 (Lorenz Breidenbach, 15 April 2021 #51) articulates a vision of leveraging the concept of introducing a decentralized Oracle network to act as a abstraction layer between decentralized nodes and interfaces for smart contracts. Those contracts are programmed and cryptographically secured, creating the possibility of using computing resources based on set contractual obligations.

An example: A smart contract based on blockchain technology (Lorenz Breidenbach, 15 April 2021 #51) can be programmed to exchange a certain attribute of PII (first name) to be used by the service (Facebook) to process a contract (username for posting a comment). This mechanism can be rolled out for any service; currently it is mostly used for financial transactions (Lorenz Breidenbach, 15 April 2021 #51). For this transaction to work, Facebook would require an interface that operates within the Oracle network as can be seen in Figure 5. Oracle network

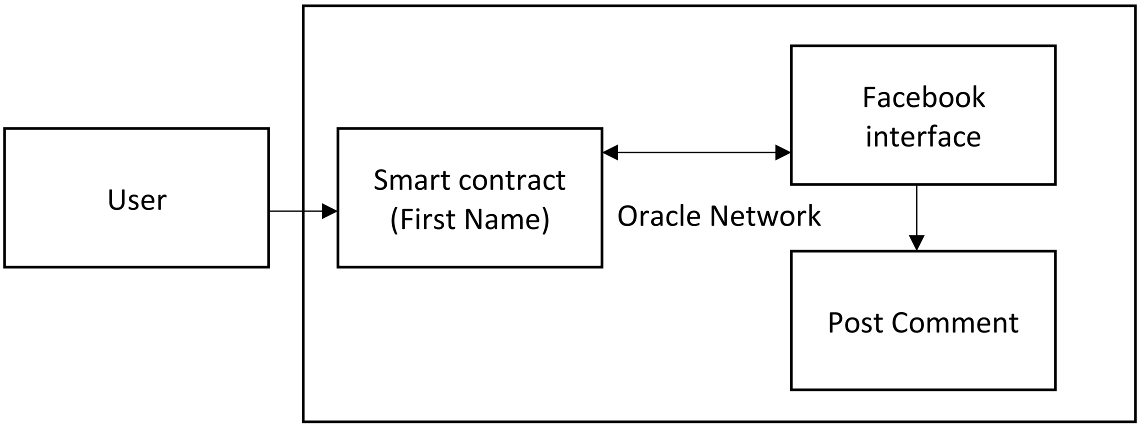

Figure 5. Oracle network

iDEAL

iDEAL is an example of a system that leverages both centralized and decentralized models to establish trust between different parties to facilitate a transaction. iDEAL is an internet payment method facilitating payments in the Dutch market. The largest Dutch banks have developed iDEAL as a standard, which allows consumers to pay online in real-time to merchants integrated with the iDEAL platform. Every consumer who has access to an internet banking product from a Dutch bank affiliated with iDEAL can send transactions via iDEAL to integrated merchants (Ideal, 01-01-2018).

Four parties are involved in an iDEAL transaction as seen in figure 6. Ideal transaction: The **consumer** who (on the internet) buys a product or purchases a service. This iDEAL payment is usually accepted by an online retailer (**merchant**). The **consumer** has a relationship with the bank where an iDEAL transaction occurs via the consumer's internet banking environment. The consumer's bank facilitates the transaction via the iDEAL platform (**issuer**). The merchant has previously concluded a contract with the bank to be able to accept iDEAL payments. The bank of the merchant is referred to as the **acquirer**. the **acquirer** facilitates messages between the merchant and the acquirer.

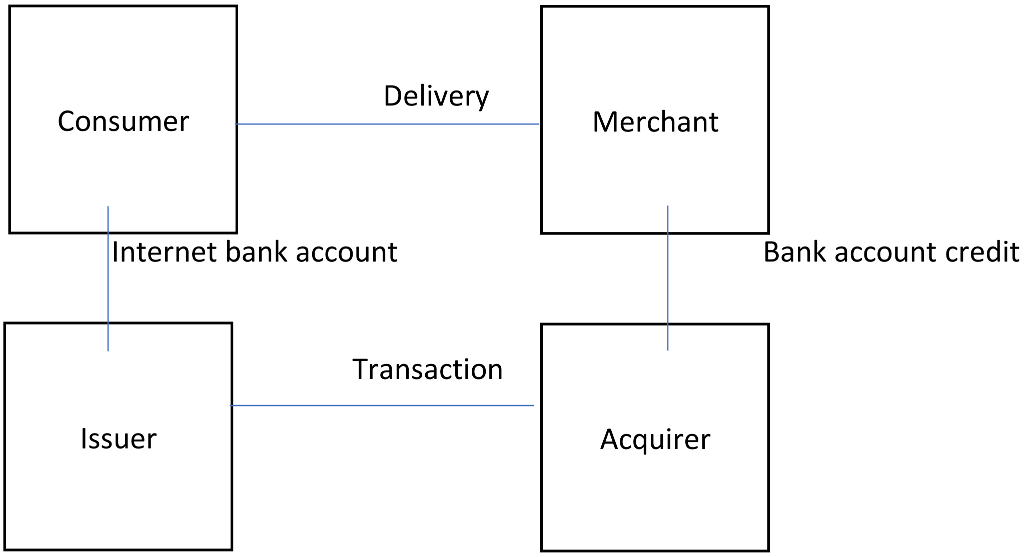

Figure 6. Ideal transaction (Ideal, 01-01-2018)

GDPR compliancy through middleware

The GDPR introduces new rights for data subjects, which are intended to rebalance power between citizens of the EU and large global companies (Auwermeulen, 2017). The GDPR introduces the concept that a natural person's privacy should be protected with respect to the data processing of their personal data by using (mostly digital) safety measures to ensure data security (Paul De Hert a, 2018).

Using a middleware solution would return power to the data subject by controlling the data the subject wants to share with the data processor. In theory, this could also alleviate the burden on the service provider to explain what the data is used for since this is a direct transaction between the data subject and data processor (Damien Geradin, 12 February 2013). Consequently, introducing middleware could potentially change the current roles of data processors and data controllers. The data subject could be their own data controller with complete control over which data processors access their data by using a decentralized middleware solution. A blockchain-based decentralized system would make GDPR audits completely transparent because all the data requested and processed by the data processor would be in the ledger (Triantafyllidis, 2016). This type of system would be a considerable improvement over the use of privacy statements, which are often not understood nor completely transparent for the data subject.

Introducing middleware also has the potential to simplify PII management since much of the burden for doing so is removed from the data processor. Nonetheless, all the security concerns related to the proper handling of PII will remain valid. In addition, centralized systems are currently difficult to audit. Many GDPR-related fines have already been given to various companies across Europe (Herman, 2020).

Conclusion

With a single authoritative system, a third party (more often than not one of the Big Five) is in control of maintaining and controlling PII. These companies use PII as a resource and profit from the current situation. However, a noticeable trend has emerged in the regulation undertaken by the European Commission. Due to increasing concerns regarding the growing power of tech companies in the European market, regulations have been tailored towards privacy-centric situations in which data ownership moves to the data subject.

A connection exists between the ownership of data and the effort of the European Commission to create more competition in the tech market. Since the EU is not directly control any of the large tech companies, regulations have been implemented to create a competitive edge and give European companies the opportunity to compete with the Big Five. The large number of lawsuits are proof of this.

Due to the change in regulations, technologies and frameworks are being developed to facilitate the move in data ownership from the data controller to the data subject. Centralized and decentralized systems are being developed to address PII management globally. To this end, middleware solutions can facilitate the proper control and ownership of PII and help data processors implement the GDPR. There are significant differences between centralized and decentralized systems in terms of both architecture and technology but also in terms of privacy and PII management. Centrally managed systems tend to establish trust by acting as a third party to manage a subject's data. Decentralized systems give this power to the user.

Expanding on this trend is the prospect of a digital environment in which an EU citizen controls their single digital identity, hence their own data, and chooses how they would like their data to be used. As a goal to work towards, a single digital identity to authenticate and store an individual's PII can be imagined. This single identity would contain all the values (e.g., age, gender, etc.) that are required to interface with any service provider. An EU citizen would have full control of what values to share with a service provider, thereby, creating more privacy and competition.

This concept can even be expanded further, such as if this identity aggregates its own search history, for example. Consequently, an EU citizen would be able to share their search history directly with service providers thus eliminating the current monopoly of the Big Five. Citizens would be able to build their own digital profiles and choose the services with whom they would like to share that profile, rather than the Big Five owning all the subjects' data and reaping all the benefits accruing from doing so.

Going one step further, a situation can be imagined in which this identity contains every single piece of aggregated data about that user and in which a user could switch from using "Service A" for delivery to "Service B" for delivery without "Service A" possessing any of the data, and "Service B" only receiving the data needed to deliver the service. Or a user's sleeping rhythm being recorded by a watch, and the user then owning this data to aggregate it into a service that gives sleeping advice, without that service receiving any other unnecessary information such as the user's favorite color or driver's license number.

Changing PII ownership from the Big Five to the data subjects would create market competition because services would not be competing for a data monopoly but rather to deliver quality services. Quality service would be required for a competitive advantage because the data subjects choose who receives the data, not the other way around. Consequently, outlining this objective and researching what the costs and benefits would be in the future might change the business models of the big five tech companies, thus, creating the competitive environment towards which the EU is maneuvering. The possibility of creating such an environment is the subject of the empirical research conducted for this thesis.

2.3 Objective of the empirical research

The objective of the empirical research is to understand what establishing a future in which complete PII ownership was retained by the data subject would mean for EU citizens and tech companies. In this regard, it is clear that decentralized middleware frameworks have more potential then centralized systems. However, all the systems currently in use are centralized on a large scale. Both solutions will be analyzed as potential frameworks for PII middleware management.

The empirical research is largely focused on the maturity level and different dimensions of the system in order to research possible future directions. The central question is focused on the decentralized framework: *How mature does the DNS-IDM framework have to be to have added value as a middleware solution for scaling while taking into account privacy, trust, and security dimensions?*

To enrich the research and provide a clear picture, other dimensions are also examined:

- How is trust established in a decentralized transactional system?
- How can PII exchange be transparently audited in a blockchain-type ledger?
- Could blockchain-based technology be the future of PII management and remove much of the burden currently experienced by data processors and controllers?
- How is it technically possible for a data subject to control their PII and aggregate all their own data?
- What are the major obstacles in introducing a single identity on the internet?

- In what way does introducing a single identity and giving control of this identity to the data subject impact on security and privacy?
- How would a user be able to control their data? How would a user be able to retract their data from a system from the service provider?

The empirical research will be conducted using the Delphi study method to gather insights from the opinions of multiple experts in the fields of privacy, competition, and identity management and to gain insights into the development of this technology in order to move towards using different business models.

3. Methodology

3.1 Conceptual Design

The objective of this research is to outline a future regarding PII ownership. When analyzing the trends found in the literature review, a future in which EU citizens are more in control of their data can be distinguished. The empirical research focuses on answering the previously stated questions to ascertain the opportunities and barriers to such a future.

To gain in-depth qualitative knowledge in each domain (Technology, trust, security, and privacy) a Delphi study is conducted. For the Delphi study a survey is used in which the experts are asked in three rounds about their opinions regarding the future of PII ownership.

Delphi Study

In-depth interviews will be conducted with the experts in group form via the Delphi method (Diamond et al., 2014) to construct an image of the possible future of PII. The advantage of a Delphi study is that this method helps create consensus between experts in the field (Diamond et al., 2014) and is ideal for gathering opinions from a small group of experts to gain that consensus (Linstone, 1975). This method allows a structured group of experts to effectively communicate and allows the group as a whole to deal with a complex problem.

To conduct the Delphi study, the qualitative data obtained from the literature review will be processed and formatted into a survey. The experts will be asked to score the data provided by other experts in this survey. This quantitative data can then be coded and statistically analyzed.

Survey

This method has been chosen for its ability to *facilitate* the discovery of new findings (Mark Saunders 2016) and to encourage *diversity* and include many different views (Mark Saunders 2016). Furthermore, this method is used to *triangulate* data to corroborate the findings with the exploratory data and the data from the presentation (Mark Saunders 2016).

This tool is used to communicate with the experts in each Delphi round. The answers will be processed anonymously, and the results will be coded to conduct a statistical analysis into the predictions and consensus that result from the Delphi study.

Expert selection

The term “expert” has not been narrowly defined (Hasson, 2000). The participants and their commitment to completing a Delphi study are typically determined by their involvement with the research at hand. Therefore, there will be a balance between knowledge and opinions. A panel of informed individuals will be invited to participate (Hasson, 2000). These individuals are knowledgeable about the GDPR, information security, information technology, blockchain technology, governance models, and trust.

3.2 Technical Design

The Delphi method requires a considerable amount of the experts' time. to remove the need for group meetings, the findings are coded and structured into a survey, then returned via email to the experts. The survey enables the experts to reply to each other's findings and acts similarly to a focus group interview (Mark Saunders 2016). The survey responses are coded into a Likert scale where the experts anonymously rate the categories distilled from the interviews and the first round of the survey.

Structure of the Delphi study

The data will be interpreted and presented to the experts. When creating the survey for the Delphi study, any conformation bias will be taken into account when formulating questions. The Delphi study will have the following structure (Linstone, 1975):

- Participants will have the opportunity to respond in **three rounds**;
- **Round 1**: Open-ended questions are asked in the survey to generate ideas and request at least six opinions (Hasson, 2000). The survey will be used as a communication tool between the experts.
- **Round 2:** The results of Round 1 are coded and statistically analyzed to identify items for which there is a collective opinion until consensus is achieved (Hasson, 2000).
- **Round 3:** Issues are grouped together just as in Round 2 and coded and analyzed.
- Input is not to be changed, and the core of each item should be left as intact as possible.
- Participants will be contacted via email.
- The results of each round are shared with the experts anonymously.
- The feedback mechanism: The information that is obtained will be shared with the experts at predefined times for three rounds. Each report will represent the input for the next round of discussions. This can be done directly and without reflection. Based on the answers, new or more detailed questions can be developed and presented to the participants in the next round. In this way, an attempt is made to begin a conversation on a matter for which the experts do not directly know the solution. The fact that the participants are able to provide feedback on each other's views may lead to additional variation with regard to the possible solutions to the problem. The number of supporters for each opinion is made available to the experts to induce convergence.
- All the research questions will constitute the input for the discussion. Any different biases and interests will be accounted for. The aim is to recognize different points of view, which supports better cooperation between the different experts.
- There will be room for intervention between each rounds.
- The statistical findings are coded.
- The experts remain anonymous to each other.
- The aim is to achieve a minimal **50%** expert **consensus** (Linstone, 1975).

Some considerations must be taken into account while conducting the research using the Delphi method (Linstone, 1975):

- If the survey is not properly and carefully curated, the identity of the experts may be revealed, and the findings may be tainted.

- The results are dependent on the chosen experts. There is a risk of randomness presented by the people involved in the study and the times at which they participate.
- Because the results are anonymous, tracing malintent or propaganda is difficult.

Experts

The experts are recruited from different fields. Because of the different aspects pertaining to PII ownership (privacy, technology, information security), experts from tech companies and professors who have **suitable work experience in the domain of this study** will be recruited**.** All the experts need to be familiar with the basic GDPR, privacy, and information security aspects. The experts are selected using a purposeful sampling for qualitative data collection. The intention is to encompass a wide range of experiences and perceptions.

The experts are selected on merit and their expertise in the subject. The experts need to be actively working or researching the fields of PII, privacy, information security, or business models concerning data. The experts are coded to remain anonymous. The contact details of the experts are known only by the researcher. Representation will be assessed by the quality of the expert. Eight experts will be selected, as studies have shown that having more than 30 experts will not improve the quality of the results (Keeney, 2011).

Survey

In the first round of the Delphi study, the questionnaire is used to collect **qualitative data through unstructured questions that seek open responses**. This information is needed initially to provide the richness of data required to **formulate subsequent focused questions or statements** (Schneider, 2012). The first round is used to explore and generate ideas. The survey questions included in the first round consist of the following:

- What are the obstacles to introducing a single identity (to be used by the consumer) on the internet?
- What challenges to establishing trust in a decentralized transactional system exist?
- Would initiatives such as Serto, Sovrin, ShoCard, Cambridge Namecoin, Blockverify, and Cambridge Blockchain, which are based on blockchain technology, be viable options for the future?
- How can technology remove much of the burden carried by data processors and controllers regarding PII governance?
- Would it be technically possible for the data subject to control their PII and aggregate all their own data?
- What are the security and privacy concerns related to introducing a single identity and giving control of this identity to the data subject?
- How would a user be able to control their data?
- How would a user be able to retract their data from a system from the service provider?
- How could PII be managed in a decentralized system?

Thematic analysis of the data is performed as a means **to synthesize the responses for each survey round**. The categories revealed from the analysis are then **grouped and listed into a Likert scale** for the next two rounds. A 5-point Likert scale is used for Rounds 2 and 3 in which the lowest scoring categories are removed (Hasson, 2000) to create a fairly narrow consensus and separate the categories into manageable numbers. The highest scoring category is retained for the next round. A minimal consensus of 75% is the aim.

3.3 Data analysis

The data is analyzed using the mixed method technique. Close-ended questions are analyzed using a descriptive method and the qualitative content analysis method is used for the open-ended questions.

Data from each round is grouped and analyzed. Items with similar issues will be grouped together with a common description. Infrequently occurring items will be omitted to maintain the manageability of the resulting list (Hasson, 2000). Each item that is carried over to the next Delphi round will be carefully considered. Items that are not well established can be omitted after consideration (Hasson, 2000).

In Round 2, the data from the items in Round 1 are analyzed and structured into statistical summaries. These will be shared with the participants to respond to for the next round. Summaries are used to show the central tendencies using the mean, median, and mode, which will help the participants to ascertain the collective opinion (Hasson, 2000).

The Delphi study aggregates different opinions from diverse experts. Because the experts do not have to meet each other, they can respond freely without the biases or pressure they may experience when face to face. The disadvantage is that the interaction of working within a group is missing (Diamond et al., 2014).

3.4 Reflection

Internal validity

The problem of self-selection exists in this method (Mark Saunders 2016). The experts who participate in the research have been selected for their expertise in the field; however, the future of PII may not be decided only by these experts. This problem is remedied by including experts from a number of different fields.

The Delphi study relies on the assumption that a larger number of people are less likely to arrive at the wrong decision than merely one expert, which assumes that there is safety in using multiple experts in a panel. Because of the feedback built into each round, the decisions are strengthened by arguments when the assumptions of the panel are challenged, which helps to improve the study’s validity (Hasson, 2000). Surveys are used to mitigate the knowledge gap of some of the experts because some are from different fields, thus, increasing the internal validity. Nonetheless, it must be stated that the impact on validity arises from the response rate of the panel (Hasson, 2000).

External validity

This research is only generalizable in Europe because it focuses on the GDPR and the trend in PII ownership. Similar legislation is being introduced in California (privacy, 2020) and other places in the world as this research is being conducted. However, those places are not included in the scope of this research. Nonetheless, the conclusions of the research may be applicable outside of the EU.This study is focused on Europe, but extrapolating the vision of PII ownership and its barriers could be applicable in any situation where data ownership is moving from data processors to the data subject (Damien Geradin, 12 February 2013).

In addition, there is the challenge of interpretation. To ensure that the expert data is correctly interpreted and coded, multiple experts will review each other. This approach leads to research in which expert consensus can be obtained in contrast to only the researcher interpreting the data (Mark Saunders 2016).

Reliability

The results are dependent on personal input from experts. Because of the different perspectives and backgrounds of the experts, multiple variations are likely to be found when attempting to reliably replicate the study. This means that to an extent there is no reliability. When, under similar conditions, the same information is given to another panel, it is not 100% certain that the outcome will be the same. This dilemma is overcome by using the criteria established for qualitative studies to confirm that the interpretations of the findings are credible. There are four major criteria for assessing interpretations: credibility (being able to trust the results); auditability(the degree the process can be reproduced); conformability (the degree experts conform), and applicability(how applicable the findings of the experts are to the research) (Hasson, 2000).

The interpretation problem will be solved because the data will be interpreted multiple times by different experts. This research is focused on a prediction, which is less reliable than a case study. The purpose of this research is to focus on a concept and vision and how to achieve it, rather than to accurately describe the details of this potential future.

To make the study as reliable as possible experts from different fields will be interviewed. The experts will also be given the opportunity to comment on each other’s findings to increase the reliability of the findings (Mark Saunders 2016).

4. Results

4.1 Delphi study

Eight experts participated in the Delphi expert panel. To ensure that different dimensions were considered, experts with expertise in different fields were invited to participate, namely, experts in the fields of privacy law and information security along with an identity specialist, a technology entrepreneur/specialist and various executives and board members of technology companies (See Appendix 2.2.2 Delphi experts Round 1).

All the experts are known to the researcher and possess the following titles: **Senior Vice President Information Security & IT, Security Officer, Senior Developer, Senior Legal Counsel** and **Chief Information Security Officer.**

Three rounds were conducted successfully conforming to the design in Chapter 3.2. The participation rate ranged between 80-100% per round. The Round 1 questions were coded into the following categories: security and privacy, technology, trust, adoption, maturity, systemic risk, compliance, and business models. The codes can be found in Appendix 2.3 Delphi Round 1: The future of PII – A single identity (Coded).

4.2.1 Trust, compliancy, and risk

The panel agreed that trust is a key concept in introducing a digital identity and that trust is one of the major obstacles to introducing a system that manages PII. *Expert #4* answering Question 3 stated:

"*Working with a decentralized transactional system (DTS) requires the trust to use it. This can be a step too far for people to solely believe another unidentifiable person online. Without a zero knowledge proof you don't know whether the data/money is tampered with. Using DTS is an open door for criminal activities"* (See Section 4.2.4 Technology for a detailed exploration of the term "zero knowledge proof").

The experts agreed in *Question 1, Round 3* that introducing a governmental authority (DigiD in the Netherlands for example) to verify certain major attributes (such as first name, last name, age) that describe a digital identity is necessary to establish trust. *Expert #6* answering *Question 3* observed, "For use cases which require verification of identity without actually identifying someone, a decentralized system might be a bit harder to trust. Because who is saying 'this is valid information?'" It can therefore be seen as problematic for service providers to be able to trust claims about a digital identity. There was debate over who this authority should be. A possible solution might be to introduce multiple authorities for different claims. *Expert #6* answering *Question 3* believed this could be an authority similar to DigiD, as discussed in 2.3 Results and conclusions.

Anonymity should be guaranteed in any model to establish trust according to *Expert #7* in *Question 9*, who stated that "Privacy by design and data minimization, enforced by regulatory means" should be imposed. The consensus of the expert panel on *Question 1 Round 2* was that just as a constitution in a democracy is meant to protect its citizens against its government, so should anonymity be guaranteed to protect the user's identity from the authority and service providers.

There was disagreement on how trust can be technically established within a PII management system. Some of the panel agreed that any system introduces a single-point-of-failure. A single-point-of-failure means that if one part of the system fails, the whole system will fail, which has a negative impact on trust and compliancy (Kirsty Lever, Madjid Merabti, & Kifayat, 2013). According to Expert #5 in Question 3, “It is important to make sure there is a minimum spread in the distribution of control over transaction validations. For example, in the case of the blockchain, if a single organization manages a large part of the network responsible for transaction validation, its control becomes centralized again.” Another part of the panel considered that a number of promising technologies (discussed in 4.2.4 Technology) exist that could solve the trust issue. This division between the experts on this issue can be observed in *Round 2 Question 6.*

4.2.2 Technology

The experts agreed in *Round 1 Question 5* that the technology in the current market is not yet mature enough to facilitate a complete digital identity within which all a subject’s data is aggregated. However, there are multiple technologies on the market that are promising in this regard.

Hashing and Encryption: Exchanging data between the service consumer and provider was a point of contestation among the panel. Current hashing algorithms are seen as a possible solution to facilitate the exchange. The panel appeared to agree that this technology is mature enough in *Question 5* of *Round 1*.

Zero knowledge proof: A problem identified by the panel was that when data is exchanged, experts agree that it needs to be ensured that there is a way to ensure that the service provider handles the data properly. Zero knowledge proof is a technology that offers a reliable way to prove claims about a digital identify to the service provider without transmitting the actual content of the claim (Wang, 2014). Zero knowledge proof combined with the blockchain could be a solution to the zero knowledge of proof problem. In *Round 1*, answering *Question 4, Expert #4* noted that retracting data is not likely to be automated in the near future.

Blockchain: The future of PII management could be facilitated by using blockchain technology. The experts did not agree what type of ledger-based technology is the best suited for this type of management. The panel did agree on the fact that none of these technologies are currently scalable or mature enough (*Round 2 Question 2*).

Decentralized oracle network: This is a network maintained by chainlink nodes, which act as an abstraction layer offering interfaces for any service connected to the network. It offers interfaces for smart contracts and extensive off-chain resources and highly efficient yet decentralized off-chain computing resources. Key features of the oracle network are the removal of complexity, scaling, confidentiality, order-fairness for transactions, minimization of required trust, and incentive-based security (Lorenz Breidenbach, Alex Coventry, Andrew Miller, Sergey Nazarov, & Zhang, 15 April 2021).

4.2.3 Security and privacy frameworks and systemic opportunity

A different dimension from which to examine the single-point-of-failure is from a information security perspective. One of the concerns raised by the panel was the perceived single-point-of-failure when introducing a digital identity that aggregates data from service providers. In *Round 1 Question 1*, *Expert #3* stated

"having a single repository for all the data of a single subject in one place is too high a risk. It is much better to have fragments in different places. But this then puts the ownership requirement in a difficult place, as a subject would need to have multiple places available to them. This gets even more difficult if availability is taken into account."

There was some disagreement between the experts on possible solutions; however, this can be addressed in both centralized and decentralized models, as discussed in Section 4.2.4.

Smart contracts using blockchain-based technology as one of the proposed solutions was not deemed secure enough by *Expert #7* in *Round 1 Question 7*: "Smart contracts on a blockchain can be hacked, so the same concerns for traditional software apply. Audits and best practices are needed. From a privacy concerns point of view, I don't have any thoughts at this moment." Some of the experts did not share this point of view on the technology and believe that different technologies are available that are secure enough or a combination of different technologies can be used, such as the zero knowledge proof model. Those technologies are addressed in 4.2.4 Technology.

The panel had considerations around the physical ownership of data. Who will own what will determine the privacy and security controls that will need to be put in place (*Round 1 Question 2)*. An open standard is necessary to enable the system to be audited. The experts agreed this is a prerequisite for proceeding with any centralized or decentralized system. In *Round 1 Question 6, Expert #4* stated, "Using frameworks/software/tools to help with this could be interesting and, in the future, perhaps a logical and standard solution. Instead of external controllers collecting/processing/protecting your PII, take matters into your own hands."

The experts agreed in *Round 2 Question 5* that a digital identity aggregating data is more of a governance problem than a technical problem. Privacy legislation will determine much of the future of the digital identity.

There are privacy concerns with all the proposed technology. In *Round 1 Question 7*, *Expert #7* stated:

"The theory and the main idea behind the blockchain is that security and privacy come out of the box. If no one singlehandedly owns the data, then no one can unilaterally adjust the data. Therefore, security is a given. Privacy might be a bit more complicated. In the principle of the blockchain every individual owns all their data and chooses what to share with every corporation."

There was no clear preference for a centralized or decentralized model in *Round 2 Question 1.* For every type of technology or model used, privacy and security concerns could be raised. In sum, technologies that prevent access to any of the identity claims such as zero knowledge of proof seem to be most promising.

According to the experts, trust is a vital component, even in a decentralized model. The experts were not in agreement regarding the decentralized model. This was reflected in *Round 2 Question 2* for which no consensus on the use of the blockchain was achieved. Some of the expert panel acknowledge the power of the blockchain to leverage smart contracts and verify claims. It is interesting to note that the less technically minded panel members perceived the blockchain had less potential.

From a security perspective, there were concerns that a decentralized model could never adhere to governance requirements to facilitate important claims such as name and birthdate according to *Expert #2* in *Question 7: "* Smart contracts on a blockchain can get hacked, so the same concerns for traditional software apply. Audits and best practices are needed." Although this model is technically feasible, a more likely scenario would be a centralized model using decentralized technology. *Expert # 5* noted in answering the same question, "The main concern would be who is able to control that data. It will suffice if the system can guarantee that data is only accessible and shareable by the data subject."

4.2.4 Maturity and business model risk

In *Round 3 Question 3*, the experts agreed that the current business models are the largest obstacle to proceeding with users aggregating their own data. Data is seen as the most powerful commodity in the tech space. Aggregating this data is at the core of those models.

The current technology available to aggregate data is still too complicated for general use according to the expert panel in *Round 2 Question 6*. Promising technologies are available, however, none compare to the user friendly technology made available by the Big Five. In addition, the current technology is not scalable enough to roll out to the public. In answering *Question 4*, *Expert #6* stated, "I think these kinds of tools are a good start. But, especially the case with blockchain, this is relatively new technology that is advancing incredibly rapidly. A tool/party like this needs to stay on top of this and utilize the best 'versions' of these technologies available." Simplifying the technology around the digital identity space could incentivize adoption.

4.2.5 Major obstacles and opportunities

Obstacles

The panel felt that there is considerable discussion regarding current frameworks to facilitate identity claim exchange. There is currently no dominant organization setting the baseline for a framework to exchange and aggregate identity information. *Expert #4* expressed the opinion regarding *Question 8* that "For now this framework is missing so an overview of your PII and stakeholders is missing. Simplify the entire process for the end-user. Technical jargon seems to have an adverse effect on adoption." The missing framework is a major obstacle, hence, the development of such a framework is vital for the future of PII aggregation and the digital identity space as shown by the expert consensus achieved in *Round 2* for *Question 7.*

Anonymity should be 100% safeguarded. Although a PII aggregating model is promising, the experts agreed in *Round 3 Question 5* that anonymity should be at the core of any model. The lack of trust, as shown in 4.2.1 Trust and compliance, acts as both an obstacle and an opportunity since it introduces the idea of leveraging the current centralized power structures and combining those with technological solutions to ensure trust in a centralized PII management system.

Opportunities

In all areas, the experts see opportunities for growth; there is a trend toward developing mitigating technologies for all the major obstacles raised. Network technologies such as Sovrin, Chainlink 2.0, and others are also maturing.

According to the expert panel, centralized PII management is also possible, although many obstacles still exist as discussed in this chapter. However, all the major concerns around trust, compliance, privacy, ownership, governance, and open standards can currently be addressed with a mix of the decentralized and centralized technologies discussed in Chapter 2.

Conclusion

Overall, the experts seemed to favor a more optimistic tone than a pessimistic one. The qualitative model in Appendix Section 2.3.2 Selective code Figure 2. shows that there are more positive outlooks for the future then negative ones. This supposition was ascertained by counting the number of positive reactions in Round 1 and comparing them to the number of pessimistic answers

5. Discussion, conclusion, and recommendations

5.1 Discussion and Reflection

This research aims to forecast the future of PII by addressing the issues and main research questions.

A decentralized model with centralized trusted authorities to verify various PII claims has the potential to be a scalable model in the future. PII could be secured using modern hashing technologies, and zero knowledge of proof – although still in its infancy – could be leveraged to exchange data between the data subject and service provider without the service provider having access to the data. Such a model would also support service providers in their efforts to be compliant with current law and grant data subjects' access to their PII.

5.1.1 Limitations

As discussed in 3.4 Reflection, conducting a Delphi study to research the future of PII involves some limitations.

Internal validity

A limitation of the Delphi method is that future developments are not always predicted correctly since the experts may demonstrate some implicit bias. To correct for this, experts from diverse fields were asked to participate. In practice, the research showed that most of the experts had a technical background, hence, they also trusted in technological solutions in the future.

As stated in 3.4 Reflection, the response rate has impact on the validity of the study. Although the response rate was extremely high (around 90%+ per round), there was still a lack of responses from certain experts in the study.

The built-in feedback mechanism of allowing the experts to review each round did not have the expected result. In 3.4 Reflection, it was noted that "decisions are strengthened by arguments when the assumptions of the panel are challenged and therefore help improve the validity." This challenging of ideas did not occur as often as expected. It is assumed that not all the experts took the time to review each round and comment on the results, which proved to be a limitation of this study.

External validity

Since this study is mainly focused on the GDPR and the impact of PII ownership, all the experts used Europe as a reference for discussing future PII management. The scope of this research is therefore limited to the future of PII within the EU.

Reliability

The interpretation problem was solved by the data being interpreted multiple times by different experts. This research is focused on a prediction with the intention of providing a future outline. Therefore, a limitation is imposed when attempting to project years ahead in time. This research is meant to function as inspiration and a conversation starter on this subject. It also aims to consolidate different fields of research and expertise to focus technological, legislative, business, and political power on this new domain.

5.2 Conclusion

PII ownership has a bright future. Currently, PII is used mostly as currency by large tech corporations. Legislation such as the GDPR returns some of the power attached to PII to the consumer (Auwermeulen, 2017). The concept of PII ownership is gaining momentum since European legislation is pushing for PII ownership to open up the PII data space to European companies (Herman, 2020). Doing so could open up new business models similarly to what has taken place in the banking sector (Arnaud, 2019).

The following examines the three dimensions of trust, security, and compliance in terms of PII middleware, while also addressing the future of PII management middleware.

The Future of PII Management

What are the main factors in the privacy domain when introducing middleware to aggregate all PII?

Technologies are on the horizon that will enable a middleware system to be built that could scale and facilitate PII ownership and introduce new business models. However, the experts state that major obstacles must still be overcome such as decentralized trust and guaranteeing anonymity at scale.

Both centralized and decentralized technologies are emerging that are extremely promising. The consensus is that it is highly likely that a hybrid model will be introduced in the future. Decentralized concepts such as smart contracts (Jamila Alsayed Kassem, 2019) and zero knowledge proof technologies based on blockchain technology are currently able to technically manage PII and allow users to aggregate their own data. The panel experts were mostly concerned with the maturity of the technology and the scalability.

There are examples of banks, such as JPMorgan (Graffeo, 2021), using decentralized technologies such as blockchain while using JPMorgan’s centralized power to leverage the best of both technologies. This represents a first step in introducing blockchain-based technologies at scale. This trend could be extrapolated to PII management. Other centralized systems such as DigiD or other government-controlled systems could be observed introducing PII management technologies to make the data available to the data subjects.

The major obstacle in introducing any changes to the current system of managing PII is the business model of the Big Five tech companies (Gergely Alpár & Bart Jacobs, 2017). Their current market dominance has stagnated progress in this space. The CIAM system has been in place for over 10 years, and it does not seem that there are any incentives for the big tech companies to change their business models aside from low legislative pressure.

Trust in middleware solution for PII management

What are the main factors in establishing trust to facilitate the complete ownership of PII by middleware?

The expert panel was divided on the topic of trust in middleware. There is a philosophical argument that by introducing middleware the power of the data is being centralized, which represents an argument both for and against trust depending on one's viewpoint.

Decentralized technologies, such as zero knowledge proof built on smart contracts, have trust built into the technology (Paul De Hert a, 2018). This is a major advancement, since these decentralized technologies are mostly open source, making it possible for the public to test the code, thus, increasing the trustworthiness of the technology (Jamila Alsayed Kassem, 2019). In contrast, centralized institutions such as the government have inherent trust. A combination of the two forms seems to be the path forward as mentioned above. Open-source transparent technology in combination with trusted institutions has the potential to instill trust in the introduction of a middleware solution for PII management.

The expert panel was concerned that by introducing middleware, a single-point-of failure is also being introduced. The middleware solution would be increasingly important; hence, trust is essential. Therefore, the proposed policy (from centralized power), transparency (in policy and technology), and technology appear promising for introducing trust into a middleware PII management solution.

Security and privacy concerns of using middleware to manage PII

What are the security concerns involved in introducing a middleware solution to manage PII and how can those be addressed?

Introducing middleware to manage some of the most important pieces of information concerning a natural person raises privacy and security concerns. One of the experts stated, "So many different 'single identity' logins these days ranging from Google to Facebook to Apple. When all are connected, if the single identity is breached, everything is breached. End-users are stupid. It's a big risk to have all under one identity since most of them will most likely do stupid things and get hacked." This sentiment was upheld by a large proportion of the panel.

To address this concern, several technologies have already been developed: IRMA is (Foundation, 2019) a very promising framework discussed in this research that enables the data subject to have complete control over their claims. It also ensures a natural person's privacy when navigating in the digital realm.

The expert panel acknowledged the potential of technology to remedy security and privacy concerns. However, there are structural problems in introducing middleware: the centralization of power and introducing a single-point-of-failure. There was no consensus on possible approaches to tackling these issues. These issues should be followed up in further research.

The impact of introducing middleware to manage PII on GDPR compliance

How can the introduction of middleware help service providers with GDPR compliance?

Introducing middleware to manage PII could increase GDPR compliance. The current system relies on policies and does not point to a specific technical solution (Tankard, July 2016). A middleware solution could not only help improve compliance and auditing, it could also introduce technical controls for enforcing GDPR compliance. Middleware would simplify the process from a compliance standpoint, which would reduce the complexity for service providers with regard to monitoring and controlling for compliance.

5.3 Recommendations for practice

To forecast a future in which the data subject aggregates their own data, some prerequisites are necessary. The expert panel stated that the identity space is substantially lacking in open standards and frameworks (*Round 1 Expert 6 Question 4)*. The recommendation is to continue research into open standards and frameworks such as the Sovrin Governance Framework Working Group, CIAM, IMRA, and ABC

The technology is not yet mature enough according to the expert panel. This research has shown that although considerable potential exists, the technology is not yet sufficiently developed. Therefore, focusing on further developing and, more importantly, simplifying the technology to make it available to a larger audience is recommended. Key technologies that have the potential for scaling are all network-based decentralized technologies such as the decentralized oracle network, DNS-IDM, and the Sovrin network.

To extend the topic to reach a wider audience, the technology also needs to be scalable, which requires further research. it is recommended to expand testing with IMRA-like systems and develop these further.

A hybrid model of decentralized technology and centralized legislative power seems to have the most potential to address the expert panel's concerns for the future. This model could extend the shift in power from the Big Five tech companies to the individual with further legislation to open data and support new business models by introducing new technologies and business models to disrupt the PII market. This means funding current projects such as Sovrin and DNS-IDM but also developing hybrid solutions in collaboration with European countries to ensure that vital services and attributes for natural European citizens are available in the network.

A key point made by the panel was awareness, that is, the lack of consumer awareness about the use of consumer data. Industry experts should simplify and explain why this awareness is necessary with public campaigns. The panel was concerned about awareness and public support. The public's recognition of the value of having control over their PII could be used as a driving force for development in this domain. Demand generation could be key to changing the current power dynamic, since the value is in the data itself. It is possible the EU may be able to transfer the public from using the big tech companies to other platforms by informing their citizens and campaigning for PII ownership and against the tech domination of the Big Five.

5.4 Recommendations for further research

Research into hybrid decentralized/centralized identity models is needed. The initiatives discussed in this research are not mature enough for massive adoption. These will be discussed below.

The oracle network is a good starting point for further research into the power of services on a network bound by smart contracts and mending the gap between centralized systems and decentralized technology.

There is currently a small team of people in two universities working on the IMRA model. However, the scale is not sufficient to extrapolate any of the findings into practice. The recommendation is to scale up research into the IMRA framework and technology.

There is a lack of research into the effect of data portability governance rules conforming to the GDPR on new PII business models. The relation between these appears to represent a gap in the literature. Further research into the relationship between the data portability governance rules and new PII business models could further strengthen the future potential of PII.

This research places considerable emphasis on the hybrid model: leveraging the power of decentralized technology and the trust and legislative power of centralized systems. This is a new area of research, currently mostly operating in the banking sector. Continuing research into these systems and from other perspectives would be extremely valuable. Such research could answer the question on how to balance the two forms of technology for PII middleware.

There are still structural problems in introducing middleware: the centralization of power and introducing a single-point-of-failure. Further research into centralizing power over PII and introducing single-points-of-failure is necessary for any PII middleware solution to be mature enough for public adoption.

Bibliography

Arnaud, B. (2019). Open banking, and what it means for European fintechs and consumers – Part 1. *The EU and Europe.*

Auwermeulen, B. V. d. (2017). How to attribute the right to data portability in Europe: A comparative analysis of legislations. *computer law & security*(review 33), 57-72.

Cellerini, E. J., & Lang, C. (2018). Cyber Liability: Data Breach in Europe. *Defense Counsel Journal, 85*(3), 1-6.

Christopher Dye, M. C. M. (2021). COVID-19 vaccination passports. *Science, 371*(6535).

Comission, E. (2017). Mergers: Commission fines Facebook €110 million for providing misleading information about WhatsApp takeover [Press release]. Retrieved from https://europa.eu/rapid/press-release_IP-17-1369_en.htm

Commision, E. (2019). Antitrust: Commission fines Google €1.49 billion for abusive practices in online advertising [Press release]

Damien Geradin, M. K. (12 February 2013). Competition Law and Personal Data : Preliminary Thoughts on a

Complex Issue. *SSRN, 1*, 14.

Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology, 67*(4), 401-409. doi:10.1016/j.jclinepi.2013.12.002

Domingo, A. I. S. E. q., Á.M (Madrid, Spain, 2018.). Digital Identity: The Current State of Affairs; BBVA Research:.

Engels, B. (11 June 2016). Data portability among online platforms. *Journal on internet regulation, 5*(2), 17.

Ferdous, M. S. P., R. A (2012). A comparative analysis of Identity Management Systems. *International Conference on High Performance Computing Simulation*.

Fett, D. K. s., R. Schmitz, G. (2017). The Web SSO Standard OpenID Connect: In-depth Formal Security Analysis and Security Guidelines. . *Computer Security Foundations Symposium*, 189–202.

Financial Stability, F. S. a. C. M. U. (2019). Banks and third party providers agree on joint efforts regarding the transition to new payment rules.

Foundation, P. B. D. (2019). IRMA in detail. Retrieved from https://privacybydesign.foundation/irma-uitleg/#waarden

Gergely Alpár, F. v. d. B., Brinda Hampiholi, & Bart Jacobs, W. L., Sietse Ringers. (2017). IRMA: practical, decentralized and privacy-friendly identity management using smartphones. *10th Workshop on Hot Topics in Privacy Enhancing Technologies (HotPETs 2017)*.

Hasson, F. K., Sinead & Mckenna, Hugh. (2000). Research guidelines for the Delphi Survey Technique. *Journal of Advanced Nursing,, 32. 1008-15*(10.1046/j.1365-2648.2000.t01-1-01567.x.).

Herman. (2020). OVERZICHT GDPR/AVG BOETES EN SCHADEVERGOEDINGEN. *Online marketing security and privecy.*

Ideal, C. (01-01-2018). iDEAL Merchant Integratie Gids. In: Currence iDEAL B.V.

Jacobs, B. (2015). De overheid als verschaffer en beschermer van digitale identiteiten. *RegelMaat*(1).

Jamila Alsayed Kassem, S. S., Hector Marco-Gisbert, Zeeshan Pervez and Keshav Dahal (2019). DNS-IdM: A Blockchain Identity Management System to Secure Personal Data Sharing in a Network. *Security Issues and Solutions of Smart Contracts in Blockchain Technology, 15*(9).

Keeney, S., Hasson, F. & McKenna, H. (2011). The Delphi technique in nursing and health research. *Oxford*.

Kirsty Lever, Madjid Merabti, & Kifayat, K. (2013). Single Points of Failure Within Systems-of-Systems. *Researchgate*.

Linstone, H., Turoff, Murray. (1975). The Delphi Method: Techniques and Applications. *Technometrics*.

Logius. (2016). *Functionele Beschrijving DigiD Logius*

Lorenz Breidenbach, C. C., Benedict Chan,, Alex Coventry, S. E., Ari Juels, Farinaz Koushanfar,, Andrew Miller, B. M., Daniel Moroz,, Sergey Nazarov, A. T., Florian Tram`er,, & Zhang, F. (15 April 2021). Chainlink 2.0: Next Steps in the Evolution of Decentralized Oracle Networks. *v1.0*.

Lueks, W., Alpár, G., Hoepman, J.-H., & Vullers, P. (2017). Fast revocation of attribute-based credentials for both users and verifiers. *Computers & Security, 67*, 308-323. doi:10.1016/j.cose.2016.11.018

Lukić, J. (2017). THE IMPACT OF BIG DATA TECHNOLOGIES ON COMPETITIVE ADVANTAGE OF COMPANIES. *Facta Universitatis, Series: Economics and Organization*, 255. doi:10.22190/FUEO1703255L

Mark Saunders , P. L., Adrian Thornhill. (2016). *Research Methods for Business Students*: Pearson.

Mukalel Bhaskaran Smithamor, S. R. (21 April 2018). TMM: Trust Management Middleware for Cloud Service Selection by Prioritization. *Springer Science+Business*(part of Springer Nature 2018).

Paul De Hert a, b., Vagelis Papakonstantinou a, Gianclaudio Malgieri a, Laurent Beslay c, Ignacio Sanchez c. (2018). The right to data portability in the GDPR: Towards user-centric interoperability of digital services. *computer law & security*, 11.

privacy, C. f. c. (2020). A Letter From Alastair Mactaggart, Founder & Chair of Californians for Consumer Privacy. Retrieved from https://www.caprivacy.org/

Rasouli, H. (2019). Proposing a conceptual framework for customer identity and access management A qualitative approach. *Global Knowledge, Memory and Communication*, 23.

SaaS: Market Intelligence, SaaS Market Growth, SaaS Pricing, Supply Market Forecasts, Category Management Insights Now Available From SpendEdge. (2019). *Business Wire U6 - ctx_ver=Z39.88-2004&ctx_enc=info%3Aofi%2Fenc%3AUTF-8&rfr_id=info%3Asid%2Fsummon.serialssolutions.com&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal&rft.genre=article&rft.atitle=SaaS%3A+Market+Intelligence%2C+SaaS+Market+Growth%2C+SaaS+Pricing%2C+Supply+Market+Forecasts%2C+Category+Management+Insights+Now+Available+From+SpendEdge&rft.jtitle=Business+Wire&rft.date=2019-02-01&rft.pub=Business+Wire%2C+Inc&rft.externalDBID=XI7&rft.externalDocID=A571982683&paramdict=en-US U7 - Newspaper Article*. Retrieved from http://openuniversiteit.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwlV1Nb9QwELVoT0hItAJEKUW-wIWm8sYbb4zEoa22QIELWQS31cQf1UptqJoUxK_gLzPj2NndCqH2Yu2OrWj13irxOPPmMSbzA5HduCeUUBejWkonrLW6FkbVxuSl06ZQVkxIq_wRiT-Vp1_Up2gR2ia1SbO4AGgPmvP_nrjdjWuMIduknb0D38NFMYCfkXUckXccb8V8BVBRov85CJqjUKRvutmfAUOV5t5hEt4frIQoecBHj5Pg9vk7rSMDTwNt10bRYNC1rJTOUM0BZfl40_zx6_XhT1icB0nWCYlXKrLZnd4oOhoK7ofy23j0QGqntTKOtZWrpZtrXa2_f5isB_tuu8VkpDG1KeUr6nF-YReme-ua7Gu1wTZKQS4I06NvQ-osNVk8_fuhGXYCsy22RY-FS7h0V_ywx36b3XPNI_aHIHzDe8T4Kur7nKbSTI95jEXE8VvAO60Z8N7nCW2-RJsntDmizQe0OaHNB7Qfs9nJdHb8Pot2F9mZUkUmx1YqI6nXFPjc5HVeqrFV3tixF8IJ7yy9o3VCA-bYWuqCOvHn4LWpPeRWPmEPgFQRTRfUk_Yp4864cVlMfFk73PcKC87KkVeYqmgFtdI77CVBOo-Gpzj0P_8Mrtt2vuRoh-0lzOfUQtvjzqqdUzaLu8tSF89ueZ1ddn_5P3rONrura7fHHgajuFSH5Bbdi0D-X_-SYBw

Schneider, D. W. a. Z. (2012). Mixed-methods research. *NURSING AND MIDWIFERY RESEARCH 4E, 14*, 22.

Sham, S. (2019). *Customer Identity.* Paper presented at the Okta Forum, Rotterdam.

Shan, S., Luo, Y., Zhou, Y., & Wei, Y. (2019). Big data analysis adaptation and enterprises' competitive advantages: the perspective of dynamic capability and resource-based theories. *Technology Analysis & Strategic Management, 31*(4), 406-420.

Sovrin. (2020). Sovrin Network. Retrieved from **https://sovrin.org/overview/**

Supervisor, E. D. P. (2014). *The interplay between data protection, competition law and consumer protection in the Digital Economy*. Brussels: Eurpean Parlement.

Tankard, C. (July 2016). What the GDPR means for businesses. *Digital Pathways*, 8.

Tim McGuire, J. M., Michael Chui, James Manyika, Michael Chui. (2012). Why Big Data is the new competitive advantage. *Ivey Business School*.

Treasury, U. (2019). Application of FinCEN's Regulations to Certain Business Models Involving Convertible Virtual Currencies. *Financial crimes enformnent network*.

Triantafyllidis, N. P. (2016). *Developing an Ethereum Blockchain Application.* (phd), University of Amsterdam,, Amsterdam.

UNION, T. E. P. A. T. C. O. T. E. (2016). *REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL*. THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION,.

Vanberg, D. A. D. (2016). The Right to Data Portability in the GDPR: What Lessons Can Be Learned from the EU Experience? Retrieved from

Vlijmen, M. v. (2016-07-22). Jouw gegevens 'liggen' verspreid over het internet. *Blog.*

Vrabec, I. v. O. H. U. (11 December 2018). Does the GDPR Enhance Consumers' Control over Personal

Data? An Analysis from a Behavioural Perspective. *Journal of Consumer Policy*.

Wang, H. W. a. F. (2014). A Survey of Noninteractive Zero Knowledge Proof System and Its Applications. *The Scientific World Journal, Volume 2014*.

Wong, J. (8 October 2018). How Portable is Portable? Exercising the GDPR's Right to

Data Portability. *UbiComp, 8*(12), 10.

Graffeo, E. (2021). "JPMorgan and Citi are using blockchain technology, and other banks are considering allowing clients to hold crypto in bank accounts, Bank of America research finds." From https://markets.businessinsider.com/news/currencies/blockchain-technology-financial-institutions-jpmorgan-bitcoin-citi-cryptocurrency-transactions-btc-2021-2.

Appendix

1.1 GDPR European motives

The GDPR states that: “The protection of the natural person in relation to the processing of personal data is a fundamental right” (Supervisor, 2014). The European commission goes on that stating that the processing of personal data should be designed to serve mankind. This literature study will not focus on the main mission of the GDPR, rather the sub versed mission of leveraging this privacy policy to change the power dynamic of tech companies within Europe and creating more competition. By analysis article -20, -18, -15 and article 4.

Article 20 of the GDPR positions the data subject as controller of its data and constitutes a case privacy enhancing technologies to allow individuals to enjoy the immaterial wealth of their personal data in the data economy (Paul De Hert a, 2018). The right to data portability was introduced to mitigate the concern that processing of data subjects data leads to a lock-in affect (Engels, 11 June 2016).

Article 18 the right to right to restrict processing. The accuracy of the personal data is contested by the data subject, for a period enabling the controller to verify the accuracy of the personal data; the processing is unlawful and the data subject opposes the erasure of the personal data and requests the restriction of their use instead (UNION, 2016). Thus, giving control to data subjects about data processing

Article 15 the data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed (UNION, 2016). Moving control from the data processer to the data subject.

Article 4 Defining the scope of what GDPR means: What PII is involves: any information relating to an identified or identifiable natural person. And what processing of PII is applicable to (UNION, 2016). Setting the stage for big companies to take away their inherent global reach and bring it back to EU natural persons.

A causality between these articles is present. The articles try to move the power that big tech companies have by using PII *back to the EU citizens by giving them control of their PII to improve market competition*. In short; **Changing ownership of PII.**

1.2 Europe competitive advantage

Before the GDPR was introduced, the European commission was already worried about competition within the European market. In this chapter two cases show that there has been a concern about competition starting as early as 2006.

In 20 March 2019 the European Commission released a statement that they had fined Google for 1.49 Billion Euro’s for breaching EU antitrust rules: “By imposing a number of restrictive clauses in contracts with third-party websites which prevented Google’s rivals from placing their search adverts on these websites (Commision, 2019).

The research into this case started in 2006 when concerns arose in Europe that publishers were prohibited from placing any search adverts from competitors on their search results pages, that led to this ruling in 2019 (Commision, 2019). During this investigation the European commission ruled on two more cases. One in June 2017 for abusing its dominance as a search engine by giving an illegal advantage to Google’s own comparison-shopping service (2.42 Billion Euro’s). And a case in

July 2018 regarding illegal Android mobile practices to strengthen the dominance of the google search engine (4.34 Billion Euro's).

The 18th of May 2017 Facebook was fined 110 Million Euro's for providing misleading information about WhatsApp takeover. (Comission, 2017). Facebook failed to disclose that it would establish automated matching between Facebook users account and WhatsApp data.

Both rulings show that the European commission is concerned about the influence of big tech companies in Europe, and the power that they leverage using Personnel Identifiable Information and their monopoly to monetize that data and impairing innovation and competition within the European market. Although the GDPR was primarily created as a privacy regulation. The GDPR was also directly address the concerns of market competition: Monopolizing data by locking in data within a tech companies' ecosphere and locking out competition (Article 20), Data subject not having control of the processing of their PII and being misled (Article 18), Monopolizing data of data subjects (Article 15), And leveraging the international character of tech companies to have power over European data subjects (Article 4).

1.3 Big data control

As seen in the cases, where Google and Facebook were fined by the European commission. The European commission seems to emphasise the usage of data and data control. (Lukić, 2017) Has concluded that data is an important resource which provides opportunities for companies to make value on basis of collected data. The possibilities of these technologies position companies with big amounts of data in a competitive advantage. Article 20 of the GDPR aims to give some of that value back to the consumer (Paul De Hert a, 2018).

Restrict access to data, and consumer lock-in by big companies further expand their monopoly on this data by taking away the convenience for the user to change service (Auwermeulen, 2017). Mandating companies to open up their products and the data used within those products to the consumer is seen to be a step towards creation more competition in the European market (Vanberg, 2016).

1.4 Business Models

On the 26th of July 2019 European Credit Sector Associations (European Banking Federation, European Association of Co-operative Banks, European Savings and Retail Banking Group) and of two third party providers (the European Third Party Providers Association and the Financial Data and Technology Association) signed a statement opening up payment information to third parties. Customer can now chose to use their financial data in other systems than the one of the bank they chose (Financial Stability, 2019) The motivations are "to foster *competition and innovation in the retail financial services market* and to increase market efficiency, and to create a more inclusive environment for the unbanked and the newly banked, without losing sight of the financial system stability and consumers' rights protection. "said (Arnaud, 2019). By making data available the European Union hopes to foster competition by creating room for new business models.
Apart from protecting privacy the European Union also tries to create new business models by leveraging GDPR article 20 and article 102 TFEU. By forcing service providers to open up their data sets to their consumers and by facilitating the import and export of their data. (Vanberg, 2016)
This might change the role of the consumer, were the consumer have a more active role in interacting with service providers. Here lie opportunities for business to develop business models involving the processing of PII of consumers. Like processing financial data of consumer to give a certain insight or manage their financials.

Due to the large amount of lawsuits with high amounts of fines that the big companies have gotten in Europe for abusing their power (Commision, 2019). It can be imagined that they might want to frustrate the transition from owning PII to processing PII.

1.5 Technologies

There are technologies being developed that enhance privacy and facilitate the data exchange between the data subject and the service provider. Third party authentication tools are used to authenticate to services using the Facebook or Google identity service. These services share and collect more data from the user then is necessary for the service to operate. (Lueks, Alpár, Hoepman, & Vullers, 2017)

Attribute-based credentials (ABC's) allows the user to only disclose a minimal set of attributes to the service that it need to operate. This would allow the data subjects to control what data is being shared with any service provider. Thus, enabling a range of scenarios from fully identifying to fully anonymous. (Lueks et al., 2017)

Okta is developing a new product: Customer identity (CIAM). This is an identity driven authentication, were users are evaluated and identified. Working with an opt-in system were users give consent to apps using their data and storing their attributes for them. In this model the customer maintains control over their data. On the roadmap there is also upstream and downstream data control were users can exercise the right to be forgotten in the GDPR. This technology will be supported by the Okta identity engine: identify authenticate enrol activate authorize. Leveraging SCIM API Webhook, data analysis and passwordless technologies. (Sham, 2019)

Qlik2Shop is a company that started in 2006 with a service: “Baas over eigen gegevens” BOEG. This was a platform for users to share specific PII with other services (Vlijmen, 2016-07-22). This company eventually grew to be Qlik2shop, a web shop that enables users to buy at different web shops and get a discount. Most probably this happened to keep the company financially viable. Showing that 10000 users wasn't enough to make an identity management financially viable.
There are currently technologies that move towards a more privacy friendly method of authentication with service providers. These technologies give a part of the **control of the PII** from the service providers – often big tech companies- **back to the data subjects.**
Different approaches have been developed showing the trend of identity management is moving towards a more holistic approach. (table 1)

Disclose of information: GDPR states that PII may only be processed if there is a reason for processing (Supervisor, 2014). Models like the ABC-model can facilitate this exchange from a user perspective.

Security: Technologies like encryption, single sign-on and multifactor authentication are moving towards a more integrated approach. Companies are already leveraging those technologies in Identity Engines to facilitate secure data exchange between the data subject and data controller.

Opt-in: The GDPR states: “Consent should be given by a clear affirmative act establishing a freely given…” (Supervisor, 2014). From a user perspective this means that the data-subject always needs to be able to opt-into a service. Technologies like CIAM in combination with webhooks and identity engines might facilitate this requirement.

Authentication: Technologies like SAML and Oauth can already be leveraged to facilitate a secure authentication layer between the data subject and the service provider.

Privacy: There are currently researchers developing IRMA. This trend shows that due to the GDPR researchers are looking into new ways to privately exchange PII between data subjects and data controllers. (Foundation, 2019)

Identity attributes: Technologies like IRMA also show it is possible to store a data subject's PII without anyone having access to it. These technologies can further developed to be able to store any data subjects attribute. (Foundation, 2019)

Data sharing: To facilitate the trend in the exchange of PII between data subject and service providers microservices can be used.

Single identity: While the SaaS market is growing ("SaaS: Market Intelligence, SaaS Market Growth, SaaS Pricing, Supply Market Forecasts, Category Management Insights Now Available From SpendEdge," 2019). This opens up opportunities for CIAM to manage and facilitate this one identity journey across multiple services. Company's like google and Facebook are already doing this (Lueks, Alpár, Hoepman, & Vullers, 2017).

Data aggregation: Identity management tools like IRMA can collects a users's PII and make the PII available to service providers when necessary using the service providers API (Foundation, 2019). This trend can be found in more services, also in the Identity engines being developed.

Taking advantage of opening up data to consumers can be used to develop technologies on the current stacks giving the ownership of data back to the data subject and impacting business model opportunities.

#	PII identity layer	Method
1	Disclosing minimal amount of information	ABC
2	Security	SSO, MFA, hashing, encryption
3	Opt-in to services	API, webhooks identity engine
4	Authentication	SAML, Oauth, API
5	Privacy	IRMA
6	Storing user values	IRMA
7	Data sharing	Oracle network
8	A single identity standard	Smart Contracts
9	Data aggregation	API

Table 1

1.6 Conceptual model

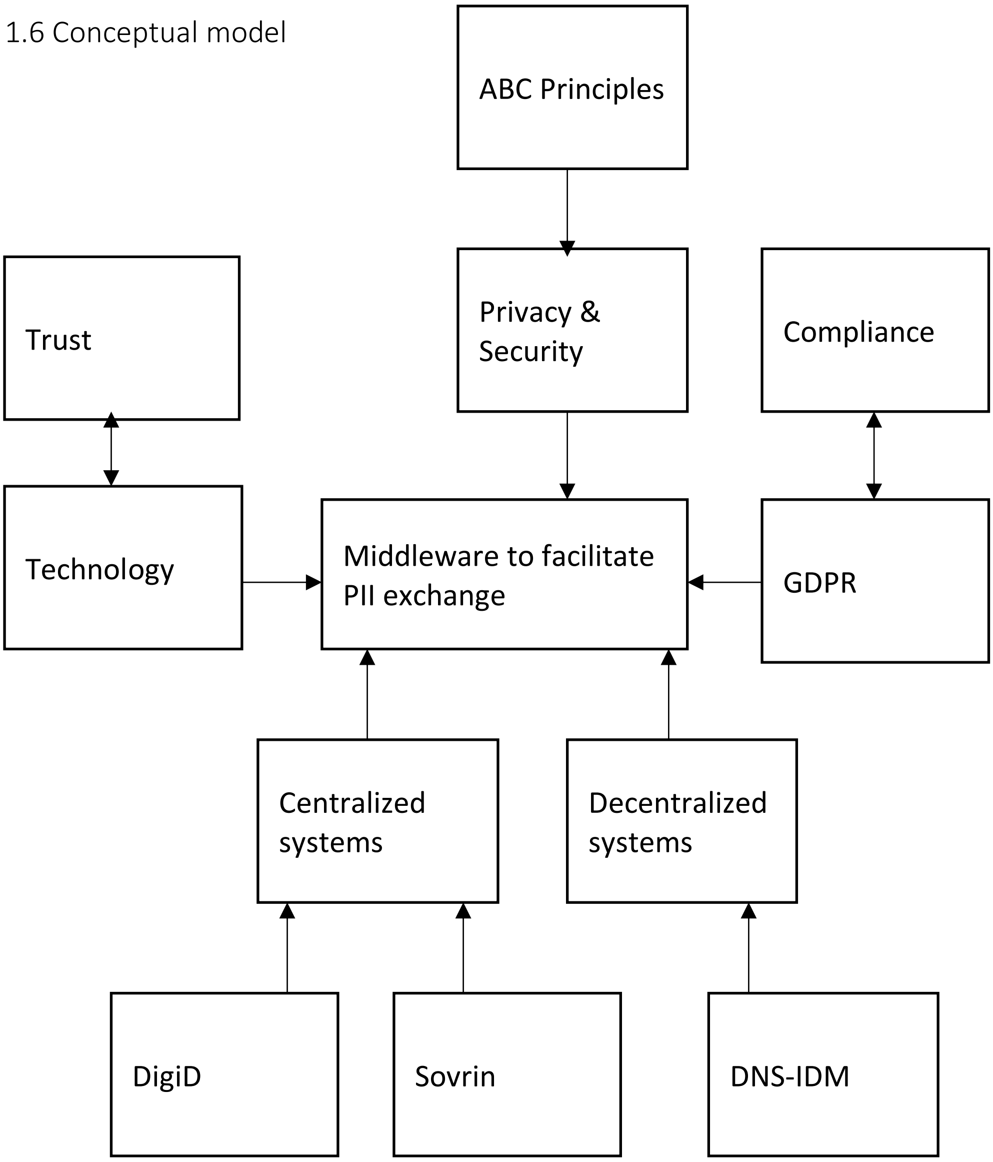

Figure. 1 Conceptual model

2.1 Delphi Study communication: Introduction letter round 1

Dear Expert,
I would like to invite you to take part in a Delphi consensus study regarding the future of personal identifiable information (PII).

This study is designed to take a look into the future of PII, and what the future of the identity on the internet might look like. The idea of having a digital identity and giving power over that data back to the data subject. In this study experts from various perspectives will partake: Legal, information security, technology and process administration
3

Why?
Taking a small part of the trend in privacy legislation; namely the effort of the European Commission to improve competitions and innovation in the European digital market. And extrapolating that to the future. And how data ownership might impact current technologies and future possibilities.

The European Parliament aims for the data economy to increase market competition to enable the development of more user-centric platforms for management of personal data. The possibility for consumers to control their own data using a middleware solution, and profit from that system, might change the internet as we know it. Just that fact is very exciting. There are many different business model opportunities for changing markets like has happened in the financial service industry by opening up financial data to third parties. This idea: of giving power back to consumers and putting them in change of the way they will interact with digital services. Owning their digital search results, address preferences etc, is the core of this study.

The focus of this study is on three aspects of a middleware systems. And how privacy, security and trust will impact the introduction and use of a middleware solution for one single digital identity management system across the consumer internet.

What will I be asked to do?
Participate in 3 rounds, the first round containing open questions, with two consecutive rounds including a questionnaire. All communication will be through email, for the questionnaire online tooling will be used. After each round, the answers of all the experts in the panel will be shared with the group.
This technique seeks to obtain a consensus opinion among an expert panel through three rounds. The responses of each round are fed back to the participants, who then in turn have the opportunity to respond to new information.
The questionnaire is expected to take around 30 minutes. In order to allow timely conclusion of the study I would request a response time of 10 days for the completion of round 1. Between every round there will be a two-week interval.
All responses will be processed and published anonymously. Your name and job function will be known by the researcher; however, the data is not tracible to any person. Any results that are shared within the Delphi panel will be shared anonymously with the group.

You will have access to the thesis once published (on the OU website). Next to the thesis a summary of results will be made available for those interested. This will be published and shared (via email) after the study has been completed.

Round 1

- Questionnaire open questions
- +/- 30 minutes

Round 2

- Responses round 1
- Questionnaire closed questions Likert scale
- +/- 20 minutes

Round 3

- Responses round 2
- Questionnaire closed questions Likert scale
- +/- 20 minutes

Thank you for participating in this panel, I look forward to learning from you.
Kind regards,

Rene van Ewijk.

2.2.1 Delphi Round 1: Design

Round 1: (in form)

Please answer the following questions

- What are the three main obstacles in introducing a single identity (to be used by the consumer) on the internet?
- What are the three challenges that exist to establish trust in a decentralized transactional system?
- Would initiatives like Serto, Sovrin, ShoCard, Camebridge Namecoin, Blockverify and Camebridge based on blockchain technology be viable options for the future?
- How can technology remove much of the burden on data processors and controllers regarding PII governance?
- Would it be technically possible for the data subject to control their PII and aggregate all their own data?
- What are the security and privacy concerns, introducing a single identity and giving control over this identity to the data subject impact?
- What would be the optimal way for a user be able to control their own data?
- What would be the optimal way for a user to retract their data from the service provider's system?
- How could PII be managed in a decentralized system?

2.2.2 Delphi Round 1 : Overview experts

Experts and fields

Candidates**	Function	Email**
anonymous	Advocaat	anonymous
anonymous	Security Officer	anonymous
anonymous	Technology specialist	anonymous
anonymous	GDPR & Contract Expert	anonymous
anonymous	Identity specialist	anonymous
anonymous	CIO	anonymous
anonymous	Enterprise architect	anonymous
anonymous	CISO	anonymous
anonymous	Entrepreneur	anonymous

**Candidates & Email known by researcher

2.3 Delphi Round 1: The future of PII – A Single Identity (Coded)

Question 1 :	*In what type of system could PII be optimally securely managed and owned by the data subject? (1,2,3,5)*	**Code**	**Axial code**
Expert #1	I don't think there is one type of system in which PII can be optimally managed. These days, PII of data subjects are located everywhere and there's absolutely not control on it. For me, step one is getting control of the PII. Ideally it would all be located in one place, for instance by an identity provider, and that third parties can read the information that they need to provide their service. However, there are so many more questions behind that, I think you need an entire paper to answer this question!	Systemic risks for PII Management	Systemic risk
Expert #2	The security level of PII data in the classical approach lies with the vendor that was trusted by the data subject. Unfortunately, with this approach the data subject often has no choice if he wants to use the services provided by the vendor. And if the data subject wants to manage the PII data, the process is typical very painful since a lot of vendors don't offer self service regarding PII data removal, etc... Leveraging blockchain with smart contracts is a very interesting concept, but the danger here is a multitude of differents companies that would implement their own proprietary version and provide it is a service. This would result in more control for the data subject to manage the PII data, but different systems that cause more confusion for the data subject. Ideally, an OASIS standard would be created just like SAML. This would pose some challenges since choosing a blockchain network is required.	Security risks for PII Management	Security risk
Expert #3	In my opinion these two requirements are never possible to both be met. To fully be able to own the data it should be in the ownership of the data subject at all times. Making sure that the subject doesn't tamper with this data at any point will be hard to monitor. Also having a single repository of all data of a single subject in one place is too high a risk. It is much better to have fragments in different places. But this then puts the ownership requirement in a difficult place, as a subject would need to have multiple places available to them. This gets even more difficult if availability is taken into account.	Security risks for PII Management	Security risk
Expert #4	A DNS-IdM	Systemic possibility for PII Management	Systemic opportunity
Expert #5	In a system where either: 1. The data subject manages his/her own PII data using fully owned and managed systems; or 2. The data subject is the sole owner of encryption details with the ability to decrypt PII data	Centralized system	Technological opportunity
Expert #6	A digital system with a user friendly UI that gives an easy overview to the user, and what data he wants to manage. This can be 'easy' data, like name and address, but potentially also more detailed data like biometrics (height, weight etc),	Security risks for PII Management	Security risk

	bankaccount numbers, social security numbers etc. Depending on the use case of the tool (if it is also to be used for more formal applications/processes) than there should be some kind of verification mechanism to verify the personal data that is entered. A potential mechanism that you could think of is an integration with DigiD for example. Last but not least: there needs to be a possibility to FULLY control your data. The tool, in my opinion, can easily be used to grant accounts (or, maybe more so 'profiles' based on your digital identity where the digital ID aks as the account), but more importantly should contain functionality to remove, edit etc your PII upon your demand.		
Expert #7	Although the concept and principle of blockchain seems ideal, the lack of widespread application of blockchain would lead me to opt for a different solution. The best "system" to optimally manage PII is a stringent regulatory scheme, as a potential improvement of the GDPR for the EU – as an example -, where the certification against that scheme would be mandatory and tied to the license to operate. To put it simply, a company would have to meet these requirements, certified by an independent third party, before it is even given the option to exist. The costs linked with such independent audits would be fuelled by the fines imposed by non compliance.	Security and compliancy risks for PII Management	Compliance risk

3

Question 2:	*What are the three main obstacles in introducing a single identity (to be used by the consumer) on the internet?*	**Code**	**Axial code**
Expert #1	Soo many different "single identity" logins these days ranging from google to Facebook to apple, , If all are connected, if the single identity is breached, everything is breached, end-users are stupid, I big risk to have all under one since most of them will most likely do stupid things and get hacked.	Systemic risk: single-point-of-failure	Systemic risk
Expert #2	- Companies with low maturity are often provided with fake account information due to trust issues from their users - The choice of when identity is needed. Think of marketing forms requesting your email address for signing up to a newsletter - The clearance level of a digital single identity versus governmental issued identities. Think of digital banks that still need to perform more checks on your passport information etc...	Trust and maturity risk	Trust risk
Expert #3	- Interoperability: each system that makes a single identity available needs to be supported by every service provider possible. We already see that many service providers only support a handful of the currently available identity providers. This severly limits the possibilities of which identity providers can be used. If a consumer picks one that is not widely supported they might not be able to use certain services. - Consumer awareness: Consumers not only need to realize	Technology adoption and trust risk	Trust risk

	the dangers involved with digital identities, but also how to prevent these. Even with a single identity and their personal data in their own management, it could only take a single mistake to put themselves at risk. The identity providers can build in a lot of protection but this will always have to be measured against usability. - Trust: Instead of sharing specific data with different service providers, consumers will now share all data with one single identity provider. The impact of a security incident is therefore much higher. Consumers will probably require a higher level of trust in these services before sharing this information.		
Expert #4	Information is traceable to the same identiy thus same person. You lose your anonimity. (Sensitive)Information can be obtained, shared, damaged and misused. Your identity can be stolen	Privacy risk	Privacy risk
Expert #5	1. Getting all involved stakeholders to trust and join the initiative. 2. Complying with the many different legislations across the world. 3. Making sure that the technology used is future proof.	Privacy and security risk	Privacy risk
Expert #6	Cooperation of third parties. Though I can see easy use cases in which this way of managing ID's is beneficial for the recipient as well, as this might simplify identification and authentication processes for them, there is also a great deal of businesses whose business model revolves around the (somewhat unlimited/unrestricted) use of PII (e.g. social media). And, it is exactly the latter category that people want that full control over. Identifying relevant data. PII like names and date of birth is easy to identify and with that govern. However, other types of data might be personal data based on the context. An example could be the membership to an organization of a person. This fact alone does not have to be personal data as it does not have to directly or indirectly say something about that person's identity. However, if that organization is a church or an organization with political affiliation, that could indirectly make that same type of data personal data. I would argue that a tool that gives full ownership of you personal data should do exact that: give FULL ownership. If there are still categories left out of scope that could still lead back to me, I might not see the benefit. Technical application: though in some use cases you could argue that the recipient of the PII never has to know who he's dealing with, as long he knows it is a legit person. Instead of actually sharing PII, a unique identifyer could be used. This way the reciepient never really holds any of your PII. However, this is not possible for all use cases. In some cases catergories of PII need to be visible to the recipient. In these cases, I could see problems with ensuring that the reciepient does not somehow make copies of this for own use. (could	Privacy security and technical risk	Privacy risk

	be fact with regulation/legislation, but might not be super effective).		
Expert #7	First and foremost, profit. A single identify, owned by the individual (in a blockchain application) would mean that companies like facebook would be unable to profit from that identity. Secondly, lack of awareness and potentially trust towards that single identify setup. When people don't understand how a system like blockchain works, they will be hesitant to offer their buy-in. Finally it would be very difficult to adhere to local (regulatory) requirements when the personal data reside everywhere and nowhere at the same time.	Risk to business model big tech	Business model big tech risk

Question 3:	*What are the three challenges that exist to establish trust in a decentralised transactional system? (6,7)*	**Code**	**Axial code**
Expert #1	Too many logins resulting in users writing down passwords and being hacked. Users using simple passwords that are easily brute forced,	Security risk	Security risk
Expert #2	*I'm not sure if I get the question here* - The dependency graph of parties you implicitly trust by collaborating with one vendor. E.g. when Solarwinds was hacked, a lot of service providers had issues not even knowing they had a dependency in their application infrastructure that relied on this vendor. - Companies reselling parts of your identity data should come forward, or could still secretly sell your identity data	Trust risk	Trust risk
Expert #3	-Tampering of information: Making sure that neither the data subject, the hosting party, or the service provider has tampered with the data -Correctness of data: If the data subject revokes access to service provider it might not be able to update the data. So this might be outdated. -Impact of breach: Due to all data being in one place, the risk of a single breach is high	Security and impact risk	Security risk
Expert #4	Working with a decentralized transactional system (DTS), requires trust to use it, this can be a step to far for people to solely believe another unidentifiable person online. Without a Zero Knowledge Proof you don't know if the data/money is temperred with Using DTS is an open door for criminal activities	Trust risk	Trust risk
Expert #5	1. Making sure there is a minimum spread in the distribution of control over transaction validations. For example in the case of blockchain, if a single organisation manages a large part of the network responsible for transaction validation, its control becomes centralised again. 2. Given the distributed nature of the data in a decentralised system, security must be battle-tested and future proof. 3. How to give data subjects control over their data in case of immutable and distributed data.	Mitigated risk by decentralized technology	Mitigating technology
Expert #6	As mentioned above: for use cases which requires verification of identity without actually identifying someone,	Business model and privacy risks.	Security risk

	a decentralized system might be a bit harder to trust. Because, “who is saying this is valid information”. (e.g. with DigiD, you know the government ensures this). The above issue could potentially be triggered by transparency or community carried verification technologies like the application of blockchain. However, for the vast majority of the people this is complex material that might not win over the trust of these people. Mostly because people tend not to trust that what they do not know. Security: centralizing all your PD in the basket of one single tool, makes you incredibly vulnerable if this information ever gets compromised.	Mitigated risk by decentralized technology	
Expert #7	Lack of knowledge, lack of technological infrastructure buy-in (from the companies who currently profit by the lack of it), lack of integrations (today) to allow the widespread usage of that system.	Lack of knowledge	Adoption risk

Question 4:	*Would initiatives like Serto, Sovrin, ShoCard, Camebridge Namecoin, Blockverify and Camebridge based on blockchain technology be viable options for PII Management in the future? (5)*	**Code**	**Axial coding**
Expert #1	I think they can be used very well in offices where you have clear regulation around the use of this and users are trained. For personal regular use, providing everyone with one identity to use anything on the internet, the risk would be enormous I think.	Technology and maturity risk	Technological risk
Expert #2	Yes	Technological possibility	Technological opportunity
Expert #3	I’m not sure storing the data indefinitely in a single managed space like a blockchain is the way to go. The problem with one space to store all PII means a much higher impact the moment this system gets breached in some way. Especially if you include a single identity into it. This would mean if another party gets hold of that identity, they can take over completely with all risks involved with that. Another problem is that it is very difficult to remove data from a blockchain. This makes the right to be forgotten difficult to execute. And it will still not be completely managed by the data subject. The blockchain still needs to be hosted by a third party to prevent spoofing/tampering and to make sure it is always available. While this is distributed over different parties it still places the responsibility of hosting this data on someone else.	Technology risk	Technological risk
Expert #4	Once steps are taken and proof is provided that this is a trustworthy and reliable option, this can be huge!	Technological possibility	Technological opportunity
Expert #5	In terms of trust, having a distributed way of storing the data, with a shared responsibility of validating transactions seems	Technological risk	Technological risk

	like a compelling way to store and validate PII. - A shortcoming of Sovrin seems to be that, while you have full control over your credentials, you are still sharing a concrete set of credentials with external parties each time. Assuming we will not be be able to control what these external parties do with this data, in time this will result in a complete loss of your credentials. - An improvement to the Sovrin network might be to add validation to the network itself, where credentials would never be shared with external parties, but instead (expire-able) signed tokens can be generated from the network that indicate that your identity complies with the requirements of the external party. This way you could even choose to not have a common ID used for your identity, but a newly generated one for each new request of third parties, preventing multiple external parties to share identities.		
Expert #6	I think these kinds of tooling are a good start. But especially the case with blockchain: this is relatively new technology that is advancing incredibly rapidly. A tool/party like this needs to stay on top of this, and utilize the best 'versions' of these technologies available. Also, a great variety of tooling might create 'competition' that might not benefit the user, as you'd want 'one solution that fits all', instead of having parties that only accept tool X or Y, but not Z.	Technological risk	Technological risk
Expert #7	Absolutely; the difficulty is that we are not there yet in terms of maturity and readiness.	Technological risk	Technological risk

Question 5:	*(How) Can technology remove much of the burden on data processors and controllers regarding PII governance? (2,4,5)*	**Code**	**Axial code**
Expert #1	Solutions like named above, definitely take a step into the right direction. However, they also have their issues and broad use brings risks. The question for me is, are those risks bigger than the risk we are currently facing? And are there ways to reduce the risks the providers bring? Can we for instance use similar authentication to this identity like DigiD	Technological risk	Technological risk
Expert #2	The PII data usage could automatically by documented by rolling up all parties their information. This would save a lot of time for the DPOs.	Technological opportunity	Technological opportunity
Expert #3	By improving interoperability so processors can limit the amount of data they store. For example a processor that uses names and emails can, when needed, query the data directly from the identity provider. The service never knows the actual identity but has a reference specific to that service. When a user wants to remove the data, the identity provider can remove the reference to the identity. The service will only have a identity id at that point without any knowledge who belongs to that identity.	Promising mitigating technologies	Mitigating technologies

Expert #4	Automation, automation and automation, once optimized (continue proces) it can help lift the 'burden' but it would always require a human check	Promising mitigating technologies	Mitigating technologies
Expert #5	Technology can help, but it introduces limitations to the services of the controllers and processors. - For example, PII is often meant to identify data subjects (authentication/authorization). Technology can help isolate these processes completely, where controllers and processors don't need access to the PII data. - Advances in cryptography might also offer solutions that give users more control over data processed by external parties.	Promising mitigating technologies	Mitigating technologies
Expert #6	As stated earlier, some third parties might be welcoming the idea of not having to 'worry' about PII as they do not control any of it, or, in case no actual PII is shared but e.g. hashed data, they might not be processing 'PII' at all, avoidning risks that come with the processing of PII.	Technological risk	Technological risk
Expert #7	I don't necessarily agree that PII governance is a burden. The lack of properly implementing the GDPR requirements around controllership is what introduces difficulties for organizations.	Compliance challanges	Compliancy

Question 6:	*Would it be technically possible for the data subject to control their PII and aggregate all their own data? (2)*	**Code**	**Axial**
Expert #1	No. Even with this in place, when you provide the necessary data to enter a service, there is no guarantee that party treats the data well or doesn't sell it.	Technological limitation	Technological risk
Expert #2	Yes, it's a bit like we see with oAuth... you can choose what you want to expose of your data. Only the minimum information is requested. It will be up to the vendors if they are willing to degrade the experience for end-users who are not willing to provide all data. E.g. logging in with Amazon not knowing your age restricts their possibility of offering smart suggestions.	Technological opportunity	Technological opportunity
Expert #3	While it would be possible, as mentioned before having a single repository of data is not always better. The impact of any kind of security flaw is much higher. As a data subject probably has limited possibilies to make sure the information is always available, it might impact the usability of the data. Not to mention the difficulty of making sure the data has not been modified to positively impact the data subject.	Technological opportunity and technical risk	Technological opportunity
Expert #4	If requested, all PII needs to be hand over to the data subject, but controlling all PII correctly yourself is a challenge. Using Frameworks/Software/Tools to help with this can be interesting and in the future perhaps a logical and standard solution. Instead of external controllers collecting/processing/protecting your PII, take matters into your own hands	Compliance and methodology opportunity	Compliance opportunity
Expert #5	It would be unrealistic to expect that the data subject will always be in full control of all their PII data. Although it would be technically possible to give the data subject control over what data they share, after they share it is out of their control.	Technologically challenging	Technological risk

Expert #6	I think this will be incredibly hard, as indicted in the question about the obstacles. Especially preventing any kind of copying or other manipulation.	Technologically challenging	Technological risk
Expert #7	Technically possible, yes, absolutely. However there are millions of people who don't have or know how to operate a computer, let alone complicated blockchain implementations. So although this is the future, my belief is that we are not there yet..	Maturity challenge	Maturity risk

Question 7:	*What are the security and privacy concerns, introducing a single identity and giving control over this identity to the data subject impact? (7)*	**Code**	**Axial coding**
Expert #1	How do you prevent user their identity get breached? How can you make the login so secure that they cannot mess it up by being stupid (think about DigiD)? How can we trust these single identity providers? How do we know who has our personal data?	Trust, security, and privacy risk	Trust risk
Expert #2	Smart contracts on a blockchain can get hacked, so the same concerns for traditional software apply. Audits and best practices are needed. From a privacy concerns PoV, I I have any thoughts at this moment	Technological risks and procedural mitigating factors	Technological risk
Expert #3	- Access control: Making sure that only specific services can access a specific piece of data - Data correctness: Making sure that the data stored is up to date and not tampered with - Single silo of all data: Having the data in a single place increases the impact of any kind of security flaw - Ease of tracking: Not only does it store all identity information but, due to the requirement of restricting access, also every service provider that has access to any piece of information. So in case of a security breach the attacker will not only have a view of the personal data but also where this data is being used.	Technical opportunity	Technological opportunity
Expert #4	1 identity = 1 point of failure. Your identity is never fully secure if your not aware of where your data is. Missing the overview	Systemic risk	Systemic risk
Expert #5	The main concern would be who is able to control that data. It will suffice if the system can guarantee that data is only accessible and shareable by the data subject. Most people already have their identity scattered around the (non-)digital world. Think about all the organisations that require passport copies or basic information such as name, birth date, place of birth, etc. This is endless. A single identity system that allows a data subject to manage accessibility to PII would at least bring some order to the chaos.	System opportunity and technological risk	Technological opportunity
Expert #6	Please see the "identifying PII" in the obstacles question. I would deem this the most tricky obstacle with privacy impact. Also, if 'nobody has control' (decentralised), who is responsible in case of compromise?	Technological and privacy risk	Privacy risk

	As for security, looking at it from a risk perspective: technical measures can of course be taken to protect data. However, due to the incredibly sensitive nature of the data, however unlikely, the consequences could be disastrous.		
Expert #7	The theory and the main idea behind blockchain is that security and privacy come out of the box. If no one single-handily owns the data then no one can unilaterally adjust the data. Therefore security is a given. Privacy might be a bit more complicated. In the principle of blockchain every individual owns all their data and chooses what to share with every corporation. Think of the scenario where we all own 100 data fields. When applying for a specific service we might need to provide 10 of those 100. In the sake of simplicity individuals will not hand pick those 10 but give permission to the service to collect that 10. There needs to be strict control that the service will only use the 10 needed and not the 90 non needed.	System and adoption opportunity	Systemic risk

Question 8:	*What would be the optimal way for a user be able to control their own data?*	**Code**	**Axial code**
Expert #1	I am not sure, this is extremely difficult considering you have no control over what third parties do with your personal data. Ideally, you would know exactly where it resides and if a third parties shares it with another party, you should be made aware and be able to object to somehow keep control. This goes way further than the scope of a single identity provider	Technological challenges	Technological risk
Expert #2	Running out of time for today to answer this one, sorry!		
Expert #3	By not focusing on where the data is stored but by making it easier to limit the amount of data specific services need to execute their functionality. A possibility could be by making it possible by retrieving certificates stating certain facts from one service provider that can be used by another service. So for example if a service wants to make sure a user is above a certain age, the user can retrieve a certificate stating that they are above that age while not stating the actual age. The user then uploads this specific fact to the service provider, which can then verify the signature of this certificate to make sure it is valid. After verification this certificate will be removed.	Technological opportunity	Technological opportunity
Expert #4	If a framework is provided, it gives options. For now this framework is missing so an overview of your PII and stakeholders is missing. Simplify the entire proces, the moment the word 'data' is involved, people are scared.	Framework opportunity	Framework opportunity
Expert #5	Sharing user details as currently offered by Facebook or Google already is very user friendly. There is a central place where you manage your details and if you want to allow external parties access, it is a matter of selecting what you want to share. Any kind of system that allows management and sharing of credentials will most likely work in a similar way.	Technological and maturity opportunity	Technological opportunity

Expert #6	Tooling that allows them to manage their data, and most importantly, gives an easy overview of who has what kind of data, for what purposes, how long etc. Potentially linking to the applicable agreements/T&C, and giving them an easy option to revoke all consent.	Trust and consent as opportunity	Trust opportunity
Expert #7	Stop using the internet :P. On a serious note, there needs to be a combination of a decentralised storing system (for personal data) with a very intuitive and easy to use "front end" so that even the novice user will have full control and will not be lost in the small letters. This is something similar to the various cookie policy notification that exist out there. The "accept all" button is very large, visible and highlighted while the "customise" button takes the user through a million sub menus..	Opportunity for mass adoption	Adoption opportunity

Question 9:	*What would be the optimal way for a user to retract their data from the service provider's system? (6)*	**Code**	**Axial code**
Expert #1	I think this should be up to the service provider and their must be strict rules around this for service providers. GDPR is already addressing it, but not strong enough. When you decide to deactivate your account, you should get the option to retrieve all your data or/and to delete it from the service provider. This should be very well regulated and service providers not complying must be fined.	Service provider opportunity	Technological opportunity
Expert #2	Running out of time for today to answer this one, sorry!		
Expert #3	By having a setup where their data is stored only when necessary and only within the timeframe of the data being needed.	Technological opportunity	Technological opportunity
Expert #4	Sending in a request for purge or handing over the data. As GDPR rules say, this is possible and offices/data-handlers should comply	Compliancy Opportunity	Compliancy Opportunity
Expert #5	Most service providers need direct access to the user data to provide their services. This means that there never will be a guarantee for a data subject to know if data was fully removed upon request. Ideally, retracting, or making data inaccessible would also be managed centrally. However, since most service providers require to host the data themselves, they need to be informed. If there is a single identity system that keeps track of external parties accessing the data, this removal request could also be initiated by the single identity system.	Technological identity opportunity	Technological opportunity
Expert #6	When using hashed data, they would hardly be a need to revoke as no PII is in their systems in the first place. In the case visible data: this would then need to be from some kind of shared DB that you can revoke access to of this third party. (like revoking an api token). Alternatively, you could actively 'push' false/anonymised data to mask the personal data (e.g. replace someones name with 'deleted user')	Hashing to increase security	Security opportunity
Expert #7	Privacy by design and data minimisation, enforced by regulatory means. Companies should be mandated to send out quarterly(?) reports to all individuals with all their data. That would allow the individuals to select what they want to share or not.	Regulatory force as force for security	Compliancy opportunity

2.3.1 Selective code

The categories are accumulated and represented in the tables. The greener the categories are, the stronger it is represented in round 1 of the expert panel answers.

Green means the code is found more than 4 times in, orange 3 times, and orange 2 or less times represented categories. The colors are represented in the selective code.

Axial code (Risks)	**# codes**
Trust risk	6
Security risk	6
Privacy risk	4
Adoption risk	1
Maturity risk	1
Compliance risk	1
Business model big tech risk	1
Total	20

Table 2

Axial code (Opportunities)	**# codes**
Technological opportunity 13	13
Mitigating technologies 4	4
Compliance opportunity 3	3
Compliancy 1	1
Framework opportunity 1	1
Adoption opportunity 1	1
Systemic opportunity 1	1
Total	24

Table 3

2.3.2 Selective model

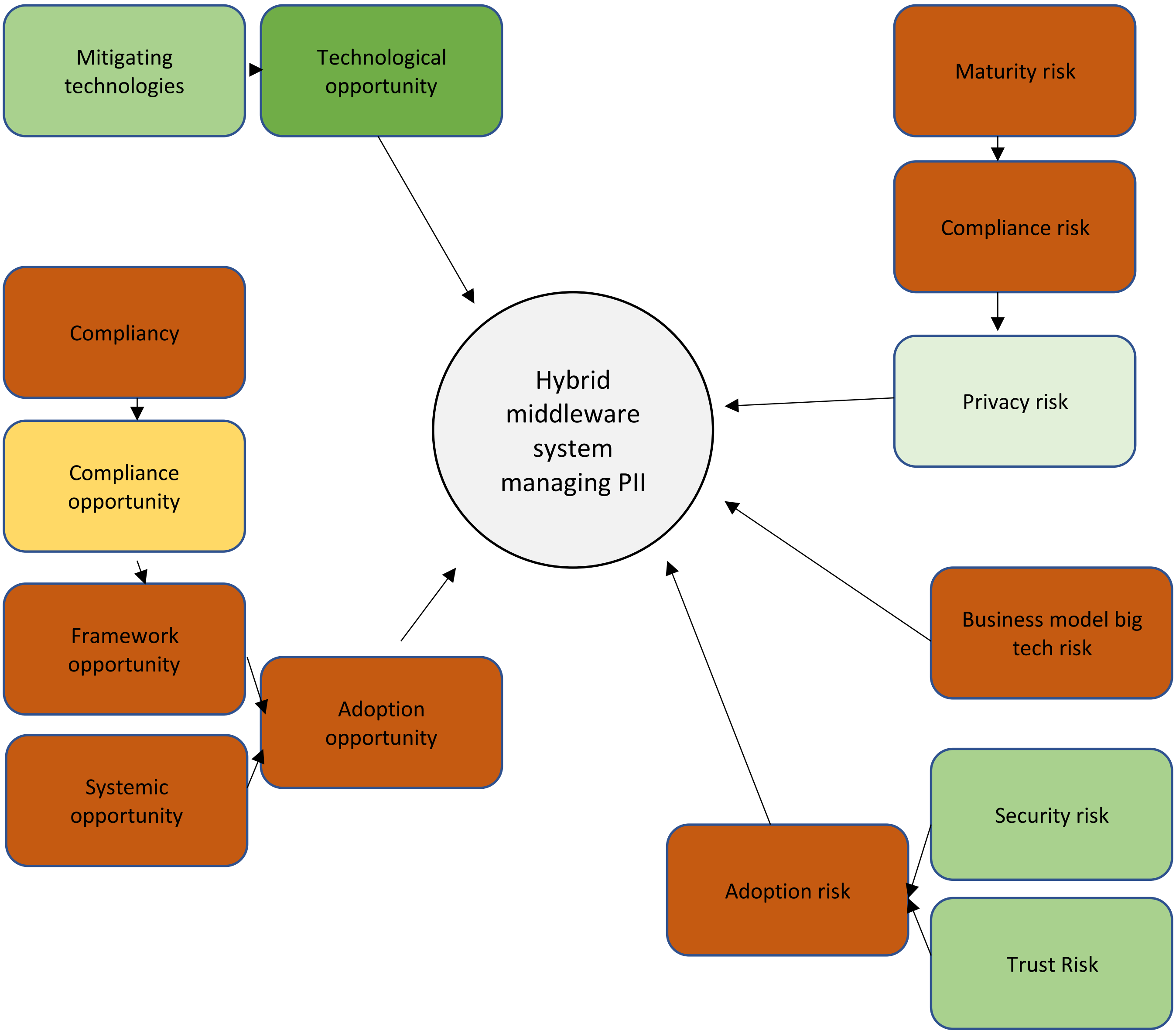

Figure 2 Selective model

2.4 Likert Scale:Study into the future of Personally identifiable information Round 2 (Responses)

The problem with even decentralised systems is that there have to be authorities governing those systems, especially on a big scale. #1	Blockchain based technologies could be the future of PII management. #2	Parts of PII management can be automated technically to remove the compliance burden on data processors. #3	Hashing PII values could be a solution for exchanging information between subject and processor. #4	PII ownership is not a technology problem. it is a governance problem. #5	It is technically possible for a data subject to aggregate and control their data. #6	Developing a framework for PII is a good next step. #7	A services that check partial hashed data for usage in the data processors system, can be the next step in PII Management. #8	PII should only be available to data processors if absolutely necessary. #9
4	3	4	3	4	2	4	2	5
3	4	2	3	3	4	5	3	5
5	5	3	4	4	2	5	5	5
3	4	4	3	3	2	4	3	5
4	5	4	4	5	4	5	2	5
5	2	4	3	4	3	4	4	5
4	4	5	4	3	2	4	5	5
4	4	2	4	5	5	4	4	2

Table 4

2.5 Study into the future of Personally identifiable information Round 3 (Responses)

If verification by an authority is an obstacle. Then PII could be verified by different authorities creating a PII profile sourced by (possibly 100s of authorities). (e.g. Google: Email, Digid: Name...) #1	Introducing new PII management software can relieve the burden of being GDPR compliancy. #2	Changing the businessmodel of Facebook and Google could be achieved by regulating PII. GDPR and the right of a data subject to upload(own) their data, to make that data available to other service providers for example. #3	Simplifying technology around PII management could incentivise data subjects to manage their own PII. #4	When centralising PII management, the possibility to stay anonymous should be build into the system. (Even if verification of an authority is necessary). #5	It is preferable, that data subjects aggregate and control their own data. #6	Zero knowledge proof offers a reliable instrument for proving the truth without revealing any other information. Can Zero knowledge proofs can help in creating a new world with a proven guarantee of trust in every transaction. Instead of partial hashing of data in transit? #7
3	5	2	5	5	3	3
4	2	4	4	4	4	4
3	4	2	5	5	4	5
3	2	3	4	4	4	5
3	2	2	3	5	4	3
4	2	2	4	5	5	3

Table 5