UNIVERSITY OF TWENTE.

Technische Universität Berlin

Consent Management on the Ethereum Blockchain

In cooperation with:

BOSCH

Submitted By:

Fabian Frank
1332368

Supervisor Bosch:
Brian Pfretzschner

Submitted To:

First Supervisor:
University of Twente
Dr. Maria-Eugenia Iacob

Second Supervisor:
University of Twente
Dr. Adina I. Aldea

External Supervisor:
Technische Universität Berlin
Dr. Maren Borkert

October 2018
Master Thesis

In this version, Chapter 5 & 6 had to be deleted to prevent sensitive information from going public before the passing of the copyright procedure.

Abstract

Exchanging personal information for access to a service has become an integral part of everyday life. Surprisingly often, we do not even realize that this exchange is taking place. When accepting the terms of service agreement of a company, it is often unclear what is happening to (personal) data. Unknowingly, users had out blank cheques to companies, allowing them to control and resell the their data. The General Data Protection Regulation (GDPR), which became enforceable in May 2018, is a first step towards putting users back in control of their personal data. With these new regulations, the existing solutions for consent management are not feasible any more for a data marketplace as well as for most consent scenarios. Utilizing the Design Science Research (DSR) methodology, this master thesis aims to create a prototype of a consent management system on the Ethereum Blockchain. With this prototype, we envision a data marketplace scenario which enables users to control their data.

Contents

List of Figures

List of Tables

"If you are not paying for it, you're not the customer; you're the product being sold" — Andrew Lewis

1 Introduction

1.1 Background

Selling personal information to a business has become part of most people's everyday life without even realising it. Trading personal user data to gain access to a service has become normal. Companies offering such a business model sell this data to third parties or use it to deliver tailored advertising. The value created through digital identities is estimated to be €1 trillion by 2020[54] which will be roughly 8% of the combined GPD of the EU-27. Due to the enormous value of personal data, The World Economic Forum has described personal data as a new asset class with an extensive ecosystem of entities collecting, analysing, and selling personal data[68]. The value of Personal Data for organizations clashes with the value that it has for the individual. Personal data has value for the individual in that it stays private while the value for the organization can only be derived through making the data more public and commercializing it. This means that exploiting the information commercially automatically means a reduction in privacy, as Acquisti et al. explain[2]. This can even lead to a decrease in overall social welfare.

In order to be able to exploit the personal information, companies have to get the individual's consent. Users agree to these conditions somewhere in the jungle of the Terms of Service, not realising the value of privacy and what they just agreed to. Most companies that utilise information technology suffer from a distinct lack of care when it comes to consent procedures [21]. Users are not adequately informed about what they are consenting to. There are examples of Software that even try to take advantage of the confusion in consent procedures (pre-ticked boxes that automatically also installs other unwanted software). Regulations have come in effect recently that aim to protect consumers from such practices. The most important one being the General Data Protection Regulation (GDPR), which is in effect since May 2018. Its objective is to put users in control of their personal data. Today third-party data trade is reliant on implicit consent, meaning that a person does not have to give specific consent to a list of companies but gives 'blank' permission. With the enforcement of the GDPR, this is no longer the case, which poses new challenges for businesses. Bosch Software Innovations experiences these challenges and have asked to find a solution which solves them. The context of the consent management system is a data marketplace scenario where a multitude of sellers and buyers can trade data. The consent management system should act as the legal structure which allows for the permissioned (re-)sale of data with explicit user consent.

The problem statement that acts as guiding theme for this thesis is: Consent is usually between two parties. In most business scenarios however, the consent that a user gives to one company serves as a 'blank cheque'. From that point on the data is traded without explicit user consent and without the user's knowledge where the data is going. The companies that collect this personal data are able to generate huge profits through the resale of given data. However, the actual

owner of the data, the individual, is left out of the further process and neither receives value from these further transactions nor insight where his data is going. New regulations are becoming enforceable which try to regulate these scenarios and make the process more clear for the individual.

1.2 Scope of this Thesis

In order to be able to comply with these regulations and make the process of consent more straightforward for consumers, this thesis aims to explore if a system utilising the Ethereum Blockchain is possible which puts the user in control of personal data. The development of such as system is explored using the Design Science Research approach by Peffers with iterative cycles for the prototype development, rigorously testing the concepts and improving on them. With this design focused thesis, this thesis aims to answer the following research question:
How can user consent be managed in a transparent and straightforward system, utilising the Ethereum Blockchain, where the user has control over what happens to their data beyond organisational boundaries?

SQ1: What does Consent Management look like today?
The goal of this question is to explore the state of the art in consent management. Looking at current mechanisms of consent will create the basis for this thesis research.

SQ2: What is the problem with Consent Management? Through this question improvement areas for consent management will be determined. By defining these improvement areas, a solid foundation for the evaluation of the developed prototype is laid.

SQ3: What is the Blockchain? (Blockchain has specific improvements for Consent Management System) Giving an introduction to the Blockchain and especially the Ethereum Blockchain will help the reader to understand why the Blockchain is an interesting architecture to try for a consent management system.

SQ4: How can the Blockchain be used for Consent Management System? This question ties in with the previous question. It is the central aspect of this master thesis as it aims to explore what a consent management system on the Ethereum Blockchain looks like

1.3 Outline

This thesis will follow the Design Science Research Methodology. The method and why it was chosen will be explained in chapter 2 after the introduction. Chapter 3 will start to identify the problem and motivation for the creation of the artefact, exploring the state of the art in consent management systems and explore the data marketplace scenario. Following on the existing solutions, current problems will be explained, and improvement areas will be identified. Functional requirements and the objective of the solution will be established

based on the issues with the state of the art in consent management along with the criteria for the evaluation phase of the different iterations of the consent management prototype.

In the following chapter, the design and development phase of the DSR will start with an introduction to the Blockchain and the different Blockchain platforms that are available will be compared. Looking at current implementations will highlight the most critical aspects, as well as possible extra functionality and explore how the application will fit in with the present application stack.

In the next chapter 5, the design phase of DSR continues. First, the basic functionality will be implemented for a consent management system on the Ethereum Blockchain. This prototype will be evaluated according to the criteria developed in previous chapters as well as consultation with colleagues from the development team of Bosch Software Innovations. The next iteration will improve on the basic functionality and will try to make the application scalable and upgradeable to achieve actual business functionality. The third and last iteration will implement all required functionality for modern consent management as identified in the systematic literature review for the back as well as the front-end. After a walk-through of the consent management process, the last prototype was also presented to a panel of potential users and evaluated using the UTAUT framework as well as the identified requirements.

Chapter 7 will take a step back and look at the broader picture, evaluating the final prototype and also examining the technical limitations of the Blockchain, the potential difficulties in a market introduction of the consent management system and other issues that may have come up during the process of the master thesis. The thesis will be finalised with a conclusion and outlook for the future.

2 Research Methodology

In this chapter, the guiding research methods are elaborated, and the choices are justified.

2.1 Design Science Research

As guiding research method for this thesis, Design Science Research is utilised. It helps guide the structure of the thesis to answer the central research questions. The methodology attempts to explore "how things ought to be in order to attain goals, and to function"[57]. With the Design Science Research methodology, the researcher's goal is to develop solutions for important problems by creating innovative artefacts that define the ideas, practices and technical capabilities in a product through which the design, implementation and use of information systems can be effectively accomplished[25]. With this process, more scientific knowledge is created, advancing the body of scientific knowledge. Hevner&Chatterjee have hypothesised that during the building of the artefact, the knowledge and understanding is created that is required to fully understand the problem at hand. Design Science Research addresses what are considered to be wicked problems[52]. That is those problems characterised by[25]:

- Unstable requirements and constraints based on ill-defined environmental contexts
- Complex interactions among subcomponents of the problem
- Inherent flexibility to change design processes as well as design artefacts (i.e., malleable processes and artefacts)
- A critical dependence upon human cognitive abilities (e.g., creativity) to produce effective solutions
- A critical dependence upon human social abilities (e.g., teamwork) to produce effective solutions

Looking at these specifications for wicked problems, DSR seems like a perfect fit for the problem at hand. Individual consent is a very ill-defined and complex environment with uncertainties on both sides, the user giving consent and the firm wanting consent. Designing a process that works for both sides requires creativity and human social abilities. Without a team that understands both sides, the researcher will most likely not find an effective solution to the problem. Hevner et al. have established 7 guidelines for Design Science Research. With these guidelines, the motivation behind using Design Science Research as guiding methodology for this thesis will be elaborated.

- Design as an artefact: Design Science Research must produce a viable artefact in the form of a construct, a model, a method, or an instantiation

In this case, the goal of the thesis process is to develop and test if it is possible and feasible to process consent management over the Ethereum Blockchain. The artefact is an instantiation, which is intended to do precisely this.

- Problem relevance: The objective of Design Science Research is to develop technology-based solutions to important and relevant business problems.

The business problem has been made clear in the introduction. Solving this problem has far-reaching implications for consent management in general and the practicability of the Blockchain in Business scenarios.

- Design evaluation: The utility, quality, and efficacy of a design artefact must be rigorously demonstrated via well-executed evaluation methods.

In academic research it is important to evaluate the created prototype using well-executed methods. In order to be able to determine whether the artefact makes sense the way it is.

- Research contributions: Effective design-science research must provide clear and verifiable contributions in the areas of the design artefact, design foundations, and/or design methodologies.

Since the goal of this master thesis, is to design an application which explores the possibilities of using a new technology for a relevant business problem, the research contribution lays a foundation for if/how a business process can be executed over the Ethereum Blockchain.

- Research rigour: Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artefact.

Utilising rigorous methods in the construction and evaluation of the artefact should be the aim of every thesis. This guideline is self-explanatory in why it makes sense to use DSR for this thesis.

- Design as a search process: The search for an effective artefact requires utilising available means to reach desired ends while satisfying laws in the problem environment.

Because of the novelty of this topic and the fact that the environment is not clear, the most effective way to get to a solution is by the process of designing the artefact.

- Communication of research: Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences.

Communicating the idea of the artefact is very important since this thesis and the artefact moves along the line of management and development.

Even though there are multiple different approaches to DSR, in this thesis the Design Science Research Methodology by Peffers et al. (2007) is chosen as an approach. The different phases can be seen in Figure 1. Peffers was chosen since it provides the most extensive and up-to-date framework for Design Science Research, including also the demonstration and communication of the artefact. This is especially important in the business context when working with developers as well as managers with less IT knowledge.

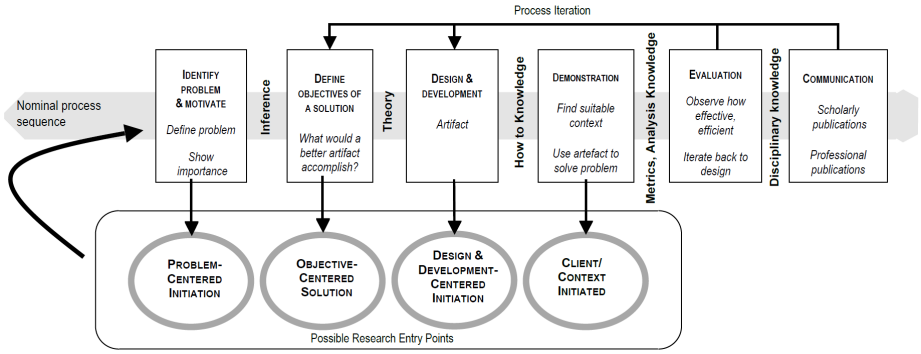

Figure 1: Design Science Research phases by Peffers et al. (2007).

Problem identification & motivation In the first phase of the DSR process by Peffers et al. (2007), the researcher aims to define the problem and justify the value of the solution. It might the useful to conceptually break the problem into smaller parts so that the solution can capture the complex problem in its entirety. The justification of the problem should motivate the reader to read on and understand the reasoning behind the problem and why the solution was chosen[47]. The researcher should be equipped with knowledge about the state of the problem and why finding a solution is important. This part of the DSR by Peffers will be done in chapter 3.

Objectives of a solution In the second phase, the researcher should determine the goal that is targeted with the solution from the identified problems and the personal knowledge of what is possible and feasible. These objectives can either be quantitative (terms that describe improvements to the current solution) or qualitative (describing how the artefact is expected to support solutions not addressed thus far). These objectives should be a rational conclusion from the problems identified in the first phase. The researcher is required to know and understand the problem as well as the current solutions (if they exist)[47]. The objectives of the solution will be determined in chapter 3.

Design & development The third phase is about the development of the artefact. The artefact can be an actual prototype, a model, method or construct. Essentially, an artefact can be any designed object where a research contribution is embedded. The activity includes the artefact's desired functionality and its architecture and then the actual creation. Important knowledge is the theory that is required to move from required objectives to design and development[47]. The design phase starts in chapter 4 with the explanation of the underlying architecture that is being used for the artefact. With chapter 5 the actual design&development phase starts.

Demonstration The demonstration phase requires the researcher to demonstrate the artefact to solve the problem. This could be an experiment, simulation, case study or other appropriate activity. The goal is to get the feedback from people who are from outside of the development team, and in

the best case actual potential users of the artefact, in order to get constructive feedback on what works and what does not. The prototype is demonstrated in chapter 5. Where the different approaches are explained and the processes are elaborated.

Evaluation In the evaluation phase, the researcher has to observe and measure how well the solution solves the problem. Here, the researcher should compare the objectives of the solution to the actual functionality of the artefact in the demonstration. This evaluation could include any appropriate empirical evidence or logical proof. The final artefact is evaluated utilising the Unified Theory of Acceptance and Use of Technology.

Communication In the last phase, the researcher has to communicate the problem and its importance, the artefact, its utility, the rigour of its design, the novelty, the effectiveness to a professional audience as well as other researchers or other audiences.

The following sections are all organised according to the 6 phases, producing multiple iterations of an artefact in the form of a software tool which solves the consent management problem.

2.2 Systematic Literature Review

A systematic literature review is one of the major tools in any academic research to support an evidence-based paradigm. The general idea is to accumulate the experiences gained from past research to arrive at the state of the art of the given topic and from that point on be able to advance the body of knowledge with a new contribution, building on the existing knowledge[56]. Such reviews follow carefully defined protocols to determine which studies are to be included, as well as for analysing their contribution in an as unbiased form as possible[13]. Budgen (2006) proposes three phase for a successful systematic literature review process. The phases are as follows: (1) planning the review, (2) conducting the review and (3) reporting the outcomes from the review.

In the planning phase, the keywords were determined which were used for the systematic literature review and the scientific databases were selected. The most important ones were Google Scholar (https://scholar.google.com) as an overall search engine as well as Scopus (https://www.scopus.com). The databases were IEEE Xplore (https://ieeexplore.ieee.org), Elsevier (https://elsevier.com), Springer (https://link.springer.com). The following search keywords were used to search the databases: "Consent" OR "Consent Management" OR "Revocation" OR "Revocation Management" OR "Informed Consent" OR "Consent Management System" OR "Electronic Consent Management" OR "Privacy". Most relevant articles were found through backwards and forward reference searching. Only English articles were included in the search and considered as credible sources and those who were no older than 10 years at the time since the internet and our interaction with the world-wide-web has changed rapidly in the past 10 years.

In the second phase, the review was conducted. The selected keywords were used to find applicable articles. Based on the found articles and their keywords, more keywords were added or keywords that seemed not important were

deleted from the search. The search was limited to the subject areas: "Computer Science", "Engineering", "Business Managmeent and Account", "Economics, Econometrics and Finance", "Psychology" and "Social Sciences". In the next stage the abstracts were scanned to determine whether the articles seemed to contain useful information based on the subject area explained in the abstract. After finding applicable papers, they were used as starting points for the backwards search on the topic "consent". One article that proved to be very good was the article "Forgetting personal data and revoking consent under the GDPR: Challenges and Proposed Solutions" by Politou et. al (2018). Since it was the most up to date article on consent at the time. For Consent Management, only Journal or Conference Papers were considered. For the Blockchain chapter, the scientific databases yielded only little results. Here, the databases were expanded to less scientific white and yellow papers due to the novelty of the topic. Only English and German papers were include in the systematic literature review. The review showed that there was a big gap when it comes to consent management and the potential use of the Blockchain. The third phase is elaborated in more detail in chapter 3.

2.3 Unified Theory of Acceptance and Use of Technology

The artefact will be evaluated using the modified Unified Theory of Acceptance and Use of Technology (UTAUT) by Venkatesh et al. (2012)[64]. The UTAUT helps to determine whether potential users of a technology artefact see value in it and does this through a collection of standardised questions that aim at different aspects of the design as well as surrounding factors. Venkatesh et al. (2012) developed the UTAUT as a comprehensive synthesis of other technology acceptance models. Previous technology acceptance models mainly focus on two aspects: Perceived usefulness and perceived ease-of-use. Venkatesh extends on these two key aspects and adds Performance Expectancy, Effort Expectancy, Social Influence, Facilitating Conditions, Hedonic Motivation, Price Value and Habit as independent variables to the model in the 2012 paper: "Consumer acceptance and use of information technology: Extending the Unified Theory of Acceptance and Use of Technology". The modified UTAUT was chosen as evaluation framework since it is tailored to the consumer technology use context.

Performance expectancy is the user's confidence in whether the technology will provide him with a benefit when performing the activity. Effort expectancy is how easy the technology seems to use to the users. Social influence describes the consumer's perceived importance that other important people in their lives think that they should be using the technology. The facilitating conditions refer to the user's perceptions of the availability of support and resources to use the technology. Hedonic motivation describes the enjoyment a user will get out of the use of the application. The price value is the perceived benefit a user gets compared to the price they have to pay. Habit describes whether it is possible that the use of the application becomes a daily/weekly/monthly habit. According to UTAUT, performance expectancy, effort expectancy, and social influence are the determinants that show the users behavioural intention to use a technology, while behavioural intention combined with the facilitating conditions determine technology use. Lastly, the individual's age, gender, and experience, are theorised to moderate various UTAUT relationships. These moderating variables are left out of the evaluation, since only a small focus

group of 5 people was used as qualitative analysis.

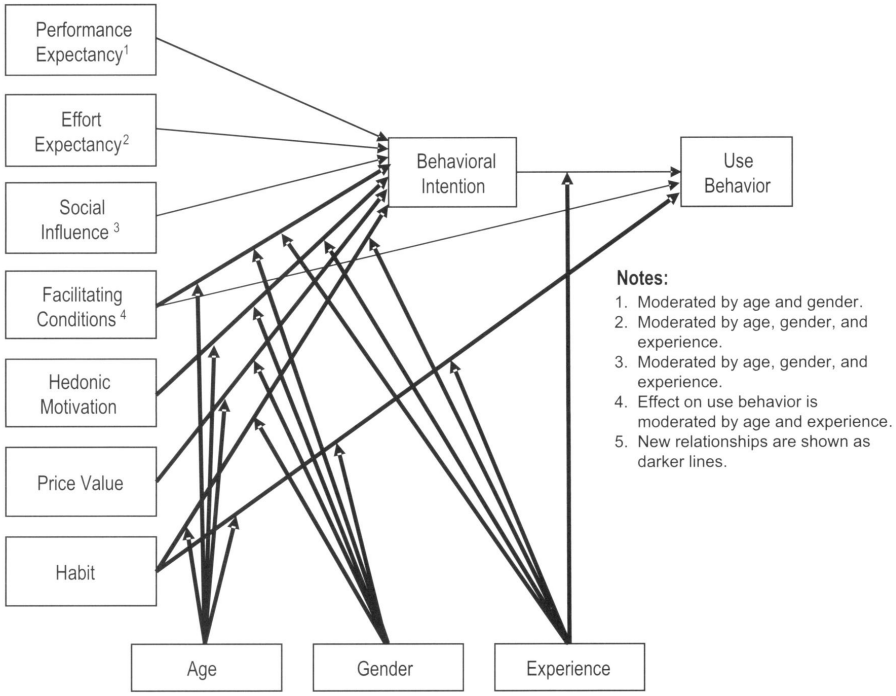

Figure 2: The Unified Theory of Acceptance and Use of Technology model by Venkatesh et al.

3 Consent Management

A consent management system allows individuals to determine what information or actions they are permitting third parties to access[46]. These systems have their origin in the healthcare sector, where the permission to access personal medical information of an individual is critical and requires extensive oversight. The concept of consent is very important since it legitimises nearly any form of collection, use or disclosure of personal data[58].

In the following section, the concept of consent will be explored and current approaches to consent management in information systems will be described be examined in order to identify problems with existing systems and translate those problems into requirements. The goal of this chapter is to identify the requirements for modern consent management. The most important requirement is that the consent process gets more transparent for the user and the communication with the individual is more clear-cut in that he is able to understand the value of his privacy and the value of personal data.

The concept of consent The Oxford Dictionaries define consent as the "permission for something to happen or agreement to do something"[45]. In a web-based context, consent usually has to be given to the terms of service of a website. This often includes the means to provide legitimate grounds for a company to collect and process user data as well as the sale of given data[48]. There are many different sorts of consent: explicit, implicit, broad, unambiguous. Each of these forms of consent are diverse in nature and need their own explanation and have been discussed in the scientific community as well as practitioners for many years when it comes to their application to research as well as the online context[59, 26, 29, 55].

The term consent is often used as synonym for informed consent. However, informed consent has a crucial distinction. Informed consent has its roots from multiple disciplines, including the medical field, law, social and behavioural sciences and moral philosophy[20] and is the "permission granted in full knowledge of the possible consequences, typically that which is given by a patient to a doctor for treatment with knowledge of the possible risks and benefits"[45]. Translating this into the context of personal data and privacy, Mont (2009) defines informed consent as a "statement that captures the willingness of individuals (data subjects) that their data could be used for specified purposes, under well-defined conditions and circumstances"[41]. The important difference between informed consent and "normal" consent is that the subject of which the consent is asked has been sufficiently informed what his data is being used for and by whom. Hereby the individual gains an actual insight into what he is giving consent to[48]. Faden&Beauchamp describe the process of informed consent as follows:
"... Action X is an informed consent by person P to intervention I if and only if[20]:

1. P receives a thorough disclosure regarding I,
2. P comprehends the disclosure,
3. P acts voluntarily in performing X,

4. P is competent to perform X, And

5. P consents to I..."

While the individual receives a disclosing of the information, whether or not the individual comprehends the disclosure is often overlooked by companies. Instead of giving the most important details in short and easy to understand sentences, consent agreements are often deliberately long and hard to understand. According to the new regulations, an individual should only have to give consent when having full disclosure over what is happening and only to a specific scenario in a well-defined time frame. Today, consent does not work like this at all. The time-frame and purpose of the consent record are usually not well-defined and leaves the company every option to sell data to whomever, whenever. Different forms of consent are often abused.

Explicit consent is a term that describes the process of giving consent with an affirmative action. This could mean to express in written or oral form that the user is willing to partake in a certain action[48]. Implicit consent, on the other hand, does not require any action and happens automatically when participating. Broad consent is the standard form of consent for most of the online big data projects for which it is impossible to determine at the point of data collection for the data will be used[39]. The secondary future uses are unknown and therefore can not be disclosed to the user.

The concept of revocation Control over data plays an important role when talking about consent and privacy, however, the actual conversion of these seemingly important topics into actual consent management systems lacks far behind. Many practitioners as well as literary scholars have argued for more user-friendly consent mechanisms and the right to withdraw consent[37, 42]. For most companies these concepts stop after the fair processing principles of giving notice and choice and the option to either opt-in or opt-out of receiving a newsletter. Not only is consent often implicit as described above but there is little to no consideration for when an individual might want to revoke his or her consent[65]. Revocation is elaborated as "the process that permits an individual to invalidate or modify previously given consent"[41]. This is an important feature of consent management, which allows the individual to, at any time, withdraw their consent to prevent the further access to the data. The balance between consent, privacy and withdrawal has been described by many researchers as a difficult and demanding task[8].

Data is often de-identified in order to protect the privacy, but when the user now revokes the right to keep the data, tracing it down in order to be able to delete all entries is challenging to say the least. The literature and practice also distinguish between the right to keep the data and the right to use the data[66]. With revoking the right to keep the data would mean that the company — in extreme cases — has to delete it from their servers. This could mean that the company has to delete the data from multiple hard drives and backups entirely. Completely removing the data seems nearly impossible because the data is shared onwards, sometimes even with other companies, copied and moved around[41]. On top of that, providing privacy friendly and auditable proof of compliance of how and when revocation was achieved is challenging[66].

Overview over given consent Another crucial feature of consent management is to provide the user with an overview of given consent. Here the user should be able to see past consent agreements and also be able to revoke given consent if not defined otherwise in the contractual agreement. This overview is crucial, since we have to consent to so many different things that losing track of what has been consented to is inevitable. In order to be able to revoke consent, first the individual has to be informed over what has been consented to. This is very difficult today, since there is not one platform where we can see every given consent. One would have to go to every individual company that one might have given consent for something and individually revoke it. This overview would allow for the individual to keep track of their personal data and be informed about the purpose and the parties that have a copy of the data, which is not the case today[41].

3.1 State of the Art in Consent Management Systems

In the following section, the most common consent mechanisms are going to be elaborated. There are a few consent management systems in place today which aim to give the user control of what data they are giving consumers access to. However, most of them lack in multiple dimensions (when it comes to the new GDPR requirements and general consent theory).

Terms and Conditions, End User Licence Agreement, Terms of Service The most common consent mechanisms are Terms and Conditions, End User Licence Agreements (EULAs) and Terms of Service (ToS). When agreeing to the End User License Agreement, the individual usually only has to click the "I Accept" button. This interaction represents the moment of consent in which the user is indicating that he/she is consenting to whatever is in the EULA, ToS or T&C[37]. Research shows that less than 1% actually pause to read what's written in these agreements[5] and that even those who should, tend not to bother. Most instantly forget this moment of consent, but they might have agreed to the on-going use of their personal data. This approach to consent and disclosure makes no attempt to see if the user has actually understood the agreement and often contains important information with deeper in meaning hidden in large chunks of text [21]. Every software comes with such an agreement and since everyone installs multiple software programs on their computer as well as utilise a wide array of different services online, one becomes numb to the information in those agreements. Companies assume (rightfully so) that users will not take the time out of their day to read this text. In order to develop a good solution that allows for informed consent, this numbness has to be overcome.

Consent Management Platforms Even though the Terms of Service or End User License Agreements are the most common form of consent, there are some new approaches to consent management. These new forms of consent management are called Consent Management Platforms (CMPs). An exemplary CMP can be seen in Figure 3. CMPs only started to surface in May 2018, along with the term Consent Management Platform, with the implementation of the GDPR. This master thesis was already in full effect at this time. CMPs

aim at obtaining consent from EU-based users to have their data processed by advertisers and marketers. Under the GDPR, there are much more stringent requirements for companies which aim at processing and selling user data. These CMPs can be used for requesting, receiving and storing user consent. These systems also make it easy for people to withdraw their consent and are transparent to third parties who rely on the user consent in order to process data. The new GDPR regulation makes the consent process a lot more complicated for a publisher who works with multiple different advertising partners and is required to obtain user consent for each individual partner.This is where CMPs come in, built on top of IAB Tech Lab's GDPR Transparency & Consent Framework, consent management platforms offer publishers a tool for more easily obtaining and managing user consent for data processing[28].

BMW CarData as example for Modern Consent Management These new approaches are technological solutions that aim to manage consent for a specific scenario. One example of those is BMW, who has introduced their CarData platform for its vehicle owners. One of the features of the platform is that the user can manage consent to allow access to car data for third parties. Here, the user can allow and revoke consent to their personal telemetric car data[11]. This approach to consent is as advanced as it gets. However, as explained in the further section, this approach also lacks some distinct advantages.

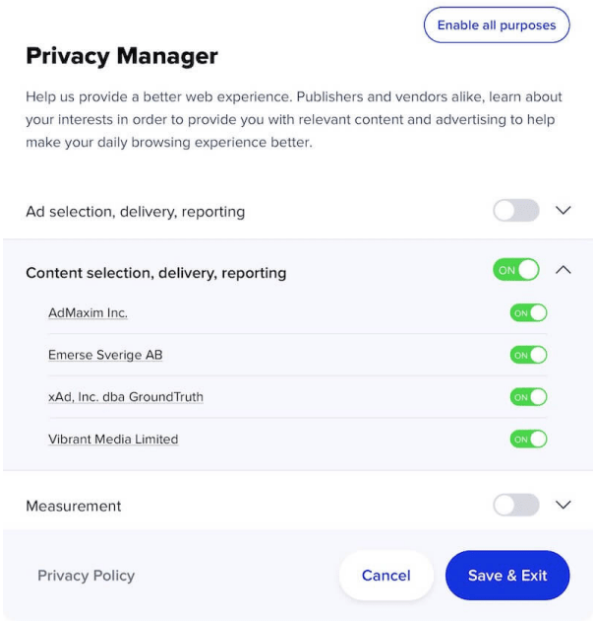

Figure 3: Exemplary Consent Management Platform

Mercedes Consent System Mercedes has also released a similar system, where they allow their users to select the released data points individually. As can be seen in the figure, the user here can select or deselect to share the specific information. This is done through a web application in the users browser. The given consent is the send to the authorization server and the web application also requests the data from Mercedes through an API.

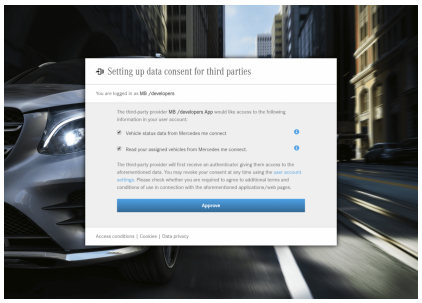

Figure 4: Mercedes Consent Management System exemplary consent request.

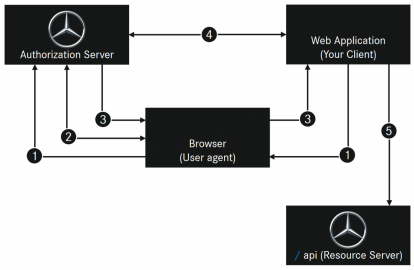

Figure 5: Mercedes Consent Management System exemplary consent flow.

3.2 Problems with Current Systems

When analysing the consent mechanism, two types of problems become apparent: one being the display and acceptance mechanism of the agreement to the user, and the other being the actual content of the consent. The display of such agreements is mostly presented as big blocks of text in small boxes, making it hard to read, while the actual content is mostly comprised of legalese and difficult terms for which you would probably need a law degree (or google every second term). Making it really hard to read and to understand the terms and conditions. One could also get the idea that these agreements are deliberately confusing and on such a high level to make it harder to read and thereby dissuade the user from reading the whole text and actually understanding it. The mechanisms of giving consent, most times just clicking on "I agree" or checking a box, are very primitive. But they would be enough when the display of the consent agreement would be improved. As it is right now, with the bad display and very basic and simple forms of agreement, the whole process just seems rushed and informed consent does not seem possible. It could give the impression that it was in the interest of the companies that the individual does not really spend time on reading their consent agreements. In order to improve the process and make the individual truly informed, either the consent mechanism or the consent display has to change[21]. Changing the consent mechanism seems to make little sense, the option of having to scroll down through the ToS in order to be able to click the "I agree" button has been explored by many companies and does not really change the fact that users do not read the text [4]. Changing the location of the button every time through randomization does also not sound like a solution that would introduce meaningful change. It would just annoy the user to have to look for the right button instead of motivating to spend time reading the agreement. A more sensible approach seems to be to change the layout of the consent agreement. Instead of having long-winded and hard to understand text, cut it down to the basics: What information is required? Who is it shared with? How long will it be retained for? What are the other important terms of the agreement?

Problem of the uninformed individual One reason why people are uninformed could be that privacy notices are long and hard to comprehend[3].

The, often deliberately, hard to understand and long winding terms of service notifications could be one reason why the individual is uninformed. Companies can hide all they want in the jungle of their ToS, always pointing the finger to the individual having had the chance to read through the whole text and be informed when complaints are made.

Problem of skewed decision-making People often lack the expertise to adequately assess the consequences of agreeing to certain present uses or disclosures of their data. People routinely turn over their data for small benefits[1]. The true value of their data is unknown to the individual, making it hard to judge whether a deal is fair or not. It is the same as having to negotiate the salary for your first job, when you do not know the value of your time, it is nearly impossible to get your true times value.

The problem of assessing harm People often favour immediate benefits even when there might be future detriments[1]. Even well-informed and rational individuals cannot appropriately self-manage their privacy due to several structural problems. There are too many entities collecting and using personal data to make it feasible for people to manage their privacy separately with each entity, since they use their own systems, if any at all, or just the ToS. Moreover, many privacy harms are the result of an aggregation of pieces of data over a period of time by different entities. It is virtually impossible for people to weigh the costs and benefits of revealing information or permitting its use or transfer without an understanding of the potential downstream uses, further limiting the effectiveness of the privacy self-management framework[58].

Due to these cognitive problems, regulators have long tried to protect the individual by establishing strong standards for how such a process should look like. In the infancy of the internet, in 1995, the Data Protective Directive was adopted. This was the first regulation aimed to protect consumers. Since then technology has transformed our lives in ways nobody could have ever imagined. Therefore, in 2016, the EU adopted the General Data Protection Regulation, which came into effect in May 2018. In the following section, the most important parts of the GDPR for consent management will be elaborated.

3.3 General Data Protection Regulation

The GDPR has come in effect in the end of may 2018. One of the most important aspects of the GDPR is that the conditions for consent have been established and now companies will no longer be allowed to use terms and conditions that are long and hard to understand and full of legalese and illegible language. According to article 7 of the GDPR regulation, the user now has the right to get the information "presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language. Any part of such a declaration which constitutes an infringement of this Regulation shall not be binding"[18]. It is also important according to the regulation that withdrawing consent is as easy as it is to give consent. Article 15 of the GDPR elaborates that the subject of the data has the right to obtain a copy of the data in the possession of the data controller[18]. Since may 2018,

users can now obtain a copy of all the data that e.g. Google has, free of charge. This includes every search term, when which service was used etc[1]. This is a dramatic shift in transparency that data collectors are required to provide. Article 17 states the right to erasure (or the right to be forgotten), which entitles the data subject to request the data holder to delete the personal data about the data subject, stop with the further trade of the data and potentially even contact third parties and have them stop the processing of the data. This is the case when the data subject withdraws consent but also when the originality intended use for the data is no longer present[18]. There are many more regulations, which this master thesis will not go into in more depth. However, failing to adhere to the regulation can lead to the company being fined $20 million or 4% of annual turnover (whichever is greater).

3.4 Conceptual Model of Informed Consent

In order to develop a consent management system which allows for a modern and more flexible approach to consent management, the literature was analysed systematically. Friedman et al. provide a conceptual model of informed consent online which is based on six components[22]: Disclosure, Comprehension, Voluntariness, Competence, Agreement and Minimal Distraction. Even though this model already exists since 2002, almost no one has applied it to a real-world consent scenario.

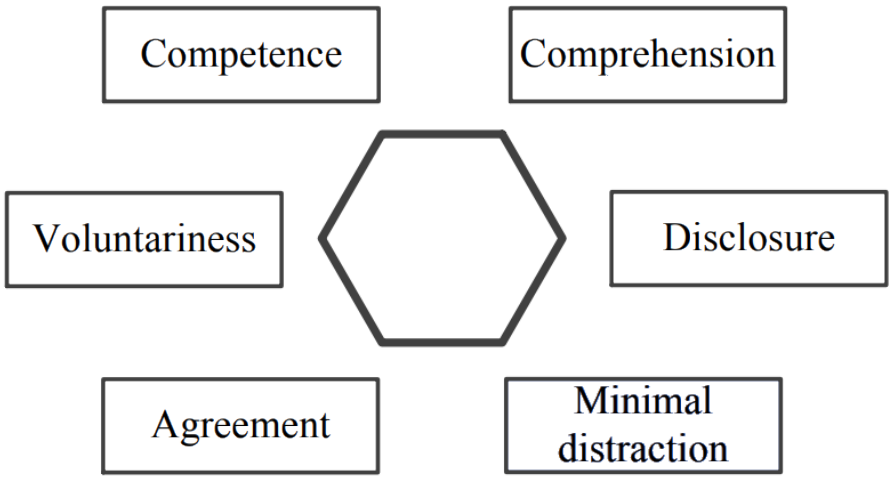

Figure 6: Friedman et al.'s conceptual model of informed consent.

The first two components, Disclosure and Comprehension are making the consent informed. The following 3 components, Voluntariness, Competence and Agreement correspond to the actual consent. Minimal distraction adds to the model that the individual should not be overloaded with information during the consent process in order to not distract from important conditions in the consent agreement. The proposed framework will be used as a guideline for the conceptual development of the consent management system.

[1]Google GDPR Takeout Tool at: https://takeout.google.com/

Disclosure Disclosure means to inform the individual about reasonable benefits and harms for the individual from performing the action under consideration. It is important that the reason or purpose for the undertaking is made clear and presented to the individual in understandable terms without too much technical detail. It is also important that any commonly held false beliefs are cleared up and that the important needs, values and interests of the user are addressed. If the action includes the collection of data from the individual it is also important that the following is made explicit[24]:

- What information will be collected?
- Who will have access to the information?
- How long will the information be archived?
- What will the information be used for?
- How will the identity of the individual be protected?

Comprehension Comprehension means that the individual has a correct understanding of what is being disclosed. However, this is hard to check for without having an actual conversation with the individual and asking questions. Friedman et al. (2005) propose two different ways, the first is being able to recite what has been disclosed in different terms and the second is being able to apply what has been disclosed to different hypothetical scenarios. A hypothetical scenario could be an e-commerce site with a recommendation system like Amazon, who recommend other products that were bought by people with similar profiles to the user. The user after getting the disclosure should be able to answer the following questions about what data is collected and how it is being used[24]:

- Will information about the customer's last three purchases be included in the recommendation system?
- Will some other user of the recommendation system be able to determine what the customer has purchased in the past?
- Will information about the customer's past purchases be a part of the recommendation system two years from now?

Without the face-to-face interaction, in technologically mediated interactions, the lack of many social and visual cues makes it more difficult to validate whether the individual has understood the disclosure. The typical online consent scenario is to click a button or tick a box that to agree to the terms of service, which are written in a text box above the button (or a click-able link to the ToS). Dialogue is almost never provided (e.g. chat) when dealing with consent online[24].

Voluntariness Voluntariness means that the individual has the choice to either partake in the action or not. This includes that the person was not overly influenced to make that choice in the form of coercion or manipulation[60].

Manipulation means that the user is intentionally influenced by someone through the alteration of the individual's perception of the existing choices. One example could be that the user is lead to believe that a certain choice has to be made in order to be able to complete the action even though it is not necessary[24]. This could mean that the user thinks that he has to give data to a company to be able to continue while this is not the case. This could be a box where consent has to be given for something that is right under the box in which one agrees to the terms of service (which has to be ticked) and make it look like the other box is also required. Or another, more extreme example, are pre-ticked boxes, where the user has to actively un-tick the box in order not consent to a specific action[48]. Another example is the manipulation of the information that a user receives, either by overloading with information or by manipulation through anxiety or fear. Friedman gives the example of a website where the user is asked so many times to agree to the cookies website that she is manipulated into selecting the accept all cookies option in order to not be bothered all the time and then fails to notice an undesirable cookie since it is hidden in the mass[24]. The third and last example of manipulation is psychological manipulation. In this scenario, a users mental process is changed intentionally by any other person. Through flattery, subliminal messages or guilt induction, a user could be manipulated to choose a specific option, even in online interactions[51].

When thinking about coercion, people often are inclined to think about extreme examples where someone is literally forcing the user to do something[24]. However, coercion can also mean that there was no reasonable other choice (e.g. buy the service with money instead of having to disclose data) than to disclose information when wanting to use a given service. In technology-mediated and online interactions this form of coercion is a serious concern since today most crucial services have moved online entirely (university applications, insurance...). Since there might be no other option but using online services, the user is coerced into this one way of conducting his business.

Competence In order to be able to make a valid consent decision, the individual has to be mentally, emotionally and physically capable[23]. That the user is competent to make these decisions on his own has to be checked by the consent seeking party[12]. A person under the age of 18 might have the technical capabilities to give consent online but might lack the emotional and mental capabilities to make a reasoned decision about providing personal information to a business on his own[22]. When designing a website online targeted at young children and adolescents, the operators have to be especially conscious about who their asking consent from and whether or not they require written consent of a parent or guardian when collecting information about their users. In the United States, this is required according to the Children's Online Privacy Protection Act (COPPA), not complying with these regulations carries a heavy fine. In the case of children and adolescents the line is relatively clear, when it comes to adults however, the lines are more blurry[65]. A grown-up with Alzheimer's or an individual with a mental disorder might not have the mental capability to determine whether giving away certain information about him/herself lies within reason. The same applies for adolescents, everyone grows up on their own pace and whether or not someone has the mental capability to

make these decisions for him/herself at 17 might be different from individual to individual.

Agreement The term agreement means that the individual has to have a clear decision to either accept or decline participation in a certain action. It has to be considered whether the agreement is ongoing by the participant and most importantly whether the ways to accept or decline participation are visible and accessible. An ongoing agreement means that the user can, at any time, withdraw consent without having to give any reason for doing so. In real life interactions, this is always the case. A participant in a research project might always just get up and leave and thereby withdraw consent to participate. In online interactions the notion of getting up and leaving is not possible, thereby consent has to be withdrawn in another way. This option is rarely provided — or considerably harder and not as straightforward as giving consent — in the online scenario. Communication in person is not permanent, what has been said is not recorded anywhere. With an online messenger service like Facebook Messenger, even though a conversation might feel as short-lived and non-permanent as a real-life conversation, it is indeed saved on Facebook's server. In an interview, Mark Zuckerberg also confirmed that Facebook has an algorithm that reads user messages and stops them from going through when conflicting with their terms of service[2]. Even though Facebook claims that Messenger data is not used for advertising purposes, recent news have shown that Facebook gives the phone numbers of their users, which they were urged to enter by Facebook in order to protect their account, to advertisers[3]. The Cambridge Analytica scandal also showed that even though users might not have clearly given consent, through the clever use of loopholes, the data of millions of users was harvested and used for political tailored advertising[4].

This example clearly shows how hard it is to withdraw consent from an online context. Often agreement does not have to be explicit and simply participating in a situation automatically equates to consent. Often when entering a situation in which we know the typical occurrences, we automatically have consented to the rules. An example is a game of football, where when entering we automatically agree to the rules of the game without having to give explicit consent. Implicit consent, in this case, has its place since the individual has disclosure and comprehension as well as competence and voluntariness, assuming that the individual was not manipulated into participation. In an online context for implicit consent to be valid, the same points have to hold up[24].

Minimal Distraction The user should not get overly distracted during the task of giving consent. This includes not flooding the user with an unnecessary high amount of information. This, to some degree, contradicts the idea of disclosure since disclosure means that the user get all information. However, what is important here is to strike a balance between providing the information that is required for the user to be informed about the disclosure, comprehen-

[2]http://time.com/money/5227844/facebook-reviews-private-messages/

[3]https://www.eff.org/deeplinks/2018/09/you-gave-facebook-your-number-security-they-used-it-ads

[4]The Cambridge Analytica Scandal at: https://www.theguardian.com/news/series/cambridge-analytica-files

sion, competence, voluntariness and agreement, but not flooding with additional unnecessary information[24].

3.5 Functional Requirements for Modern Consent Management

The contemporary articulation of consent has been stretched thin to the point of breaking[37, 36]. Consent is not clear, it's often full of illegible terms and conditions full of legalese. Users do not know or understand what they are actually giving consent to. New regulations like the General Data Protection Regulation require organizations to rethink consent and privacy when it comes to personal user data. Current consent management systems like BMW's CarData are part of the data provider's architecture. Such data providers are in charge of both, the personal user data is and the individual consent agreements corresponding to the data. The resulting centralisation of responsibilities increases the need for trust in the data provider. In addition, it is not possible for third parties to access and validate an individual user consent.

A better solution would be to divide the point of consent and data storage in order to make consent more see-through for all parties involved. Existing CMPs are focused on one scenario: Online data collection and sale to advertising firms. Another important issue with consent management is that it ignores severe human cognitive problems that impair the ability of the individual to make rational and informed decisions about the benefits and costs of disclosing their personal data [58]. To overcome these human cognitive problems, privacy notices have to become more clear-cut and the individual[41], as well as companies have to become more aware of the personal data that is traded day by day, its value and the security risks. In the past chapter, multiple issues concerning consent were elaborated. Translating these core paragraphs from the GDPR combined with the results form previous research into functional requirements. Prior to disclosing data, when being asked for consent, the individual should be informed about:[41, 38]

- What information will be collected?
- Who will have access to the information?
- How long will the information be archived?
- What will the information be used for?
- How will the identity of the individual be protected?

Not only the disclosure is important but also that it is clear and concise and easy to understand. In the best case, understanding has to be checked by asking questions that put the data into context. During the sign up for such a service, a check for age and mental illness has to take place in order to protect the individual against exploitation. The data has to be presented in a very straight-forward way without too much distraction and only through an affirmative action should consent be valid. The goal should be to have a good transaction framework with more direct information disclosure of accurate and relevant information, rather than a general full disclosure that could easily flood the user with too much information, resulting in a confused or ignorant

decision rather than an informed one[21]. It is also crucial that consent data is transparent and that there is no middleman who controls the process of consent as well as the data and can allow or disallow access. The consent data should be visible for the entire consent process chain, so that everyone can check, individually, whether consent was given. This expression of consent has to happen in an explicit way, where the individual is not coerced into agreeing to something because they do not see that a box is already checked. The action to give consent has to be affirmative and unambiguous. It should also be possible to get an overview over the consent history and to revoke consent as easily as it is to give consent. To come to a conclusion, current consent management systems need to be improved in various dimensions in order to fit to the new regulations. Consent has to be give more explicitly with the users knowledge about what they are actually consenting to. The system has to be more flexible, reliable, transparent and independent for a modern approach to consent management. Research on Consent Management Systems focuses on traditional systems[39, 49, 8, 65, 24, 35, 37, 22, 23]. There is a gap in the literature when it comes to modern approaches to consent management. In the following section, an new approach to consent management is explored.

4 The Blockchain

When thinking about a reliable and transparent architecture that has a lot to offer for the envisioned consent management system and looking at current hype topics, the Blockchain is the first thing that comes to mind. In the following section, we will explore whether the Blockchain is a good underlying architecture for such a system and which of the multitude of available Blockchains is the best fit. Practitioners are ahead of the research community when it comes to the Blockchain. Therefore, less scientific papers had to be used as sources.

4.1 Introduction to the Blockchain

The Blockchain is a technology, which was first introduced in 2008 by a white paper by the mysterious figure Satoshi Nakamoto, the identity of whom remains a mystery until this day. The idea is a fusion of multiple technical as well as economical ideas combined with cryptography and game theory. Since the identity of the inventor is still unknown, we can only take the whitepaper as source of what was the motivation behind the creation of the Blockchain. In the whitepaper Nakamoto describes that "commerce on the Internet has come to rely almost exclusively on financial institutions serving as trusted third parties to process electronic payments"[43] and explains that these transactions still have the inherent weaknesses of the trust based model. More specifically, he elaborates that merchants need more information than would be required and the third party takes higher fees due to having to mediate disputes and transactions being reversible[43]. He underlines the need for a system based on cryptographic proof instead of a third party who validates the transactions.

Many of his ideas seem to have come from the cypherpunk movement in the 1990s which was/is focused on activism that advocates for the use of strong cryptography and privacy-enhancing technologies. The main principles of the cypherpunk movement as explained by Eric Hughes "A Cypherpunk's Manifesto" are that "Privacy is necessary for an open society in the electronic age. [..] We cannot expect governments, corporations, or other large, faceless organizations to grant us privacy [..] We must defend our own privacy if we expect to have any. [..] Cypherpunks write code. We know that someone has to write software to defend privacy, and [...] we're going to write it."[27]. He further elaborates in the manifesto, that they "[..] are defending [their] privacy with cryptography, with anonymous mail forwarding systems, with digital signatures and with electronic money." [27].

"Cryptography", "digital signatures", "electronic money", when hearing these 3 terms most tech-savvy people will instantly think of the Blockchain. Assuming that Satoshi Nakamoto was part of the movement does not seem too far-fetched. Blockchains are a distributed peer to peer network that maintains a database. The special feature of this database is that once something has been written to it, it becomes immutable because the data gets saved in a block, which then gets permanently linked through cryptography to the next block. Every participating node in the network maintains a copy of the database and verifies every transaction. Through a consensus protocol the data integrity and ordering of data as well as the consistency across the geographically distributed nodes is guaranteed. Through cryptographic hash algorithms, the security of each account and its transactions are verified.The main idea of the Blockchain

as underlying technology for Bitcoin was to be able to eliminate the need for a middleman in an online transaction by solving the double-spending problem.

The double spending problem is a fundamental problem with digital forms of of payment. The basic premise is that digital money can be copied since there is only an entry in a ledger representing it[30]. Like with traditional currencies and counterfeit money, double spending leads to inflation by creating previously not existing units of the currency. This leads to the loss of trust and devaluates the currency in relation to other monetary units. To solve and counteract the double-spending problem, Satoshi Nakamoto has combined various advances and theories and created the Blockchain as an append-only log, storing transactions. All data is fully replicated across a large number of peers. Data is combined in immutable blocks which are deterministically verifiable using the Blockchain data structure. The Blockchain is fully decentralized and does not rely on a third party for trust. Immutability is achieved using hashing, which will be described in more detail later. The data is replicated across the entire network of peers, leaving everyone with the same information. Consensus is reached through a Byzantine proof algorithm like proof of work (pow), which will also be explained in more detail later. Every node participating in the network verifies every transaction. The integrity and anonymity of the network is achieved through the clever use of cryptography. In the following chapter, a short introduction into cryptography will help to understand the basic foundation of the Blockchain.

4.2 Cryptographic Foundations

To be able to understand how the Blockchain technology works, one has to take a short trip into the field of cryptography. A Blockchain is built on two very important cryptographic foundations. The most important of which are hash functions as well as public-private key encryption.

4.2.1 Hash Functions

Hash functions are the bread and butter of the Blockchain architecture. Cryptographic hash functions are mathematical trap-door functions. Easy to compute in one direction, almost impossible in the other. They allow to create a digital fingerprint of the data. The algorithm takes an arbitrary input and converts it into a fixed length output. The Keccak-256 (one kind of hash function) hash of:

"The quick brown fox jumps over the lazy dog"

is:

"4d741b6f1eb29cb2a9b9911c82f56fa8d73b04959d3d9d222895df6c0b28aa15",

when adding a single white-space at the end:

"The quick brown fox jumps over the lazy dog ",

the outcome becomes:

"75f80f0fb49a16e547d5d29e8c145a26a5aea3adda99a49e5c69b858b59ee012".

Changing even one white-space will result in a completely different outcome. One could get the idea now that the function just takes the input and randomly converts it into a fixed length output. However, this is not true. Hash functions need to satisfy multiple properties in order to be considered safe and useful for the Blockchain application. The first property is that the result of the function has to be deterministic. This means that feeding the algorithm the same data

will always result in the same outcome. If this is not the case then it is impossible to keep track of the input. One could not proof with the outcome of the hash that two inputs are identical. Another property is that the hash-function has to be pre-image resistant. It has to be infeasible to determine the input a where $H(a)$ is the output hash. The emphasis is on feasible since it is always possible to determine the input by trial-and-error. With enough time/computing power one could just feed the function with every possible input until the output hash matches to the given hash. An interesting application of these two properties can be observed in the Wikileaks publications. The organization published a hash value a on their Twitter Account of the information when they retrieve it before publishing. When actually publishing the information b, everyone can compare the hash of the document $a = H(b)$ to the previously published hash and thereby determine that nothing in the document has been changed.

Collision resistance is the next important property of a cryptographic hash function. The algorithm has to be written in a way that makes it extremely infeasible that two random inputs $H(a) = H(b)$ result in the same output $a \neq b$[53]. It is impossible to design a hash-function with arbitrary input length and fixed output length that is completely collision resistant since the input space in larger than the output space. This is known as the pigeonhole principle in mathematics which states that for m containers to put in n items, if $n > m$ then at least one container must contain more than one item[63]. The emphasis lies on infeasible. It is possible but it has to be only possible by brute-forcing in order to make it infeasible. If someone can reverse engineer the algorithm and thereby cause a collision would make the hash function useless.

The last property is uniformity. Every hash of the output range should have the same probability of occurring. That is, the inputs of a proper hash function should be mapped as evenly as possible through the output range. Collisions would be more likely if a specific output had more probability to be hit than others and this would also destroy the mechanism of mining that is used in the Blockchain. This will be explained more in detail later, but in short this is a puzzle that has to be solved by trial-and-error by the so-called "miners". The puzzle is the search for a specific value, if now the hash function had an uneven distribution, miners could change their "mining" algorithm to first look for the solution in the higher chance range. This would give them an advantage since they would be able to solve more puzzles faster and would also potentially give the opportunity to tamper with the Blockchain.

4.2.2 Public-key Cryptography

Secrecy, authenticity and integrity are the three pillars of any crypto-system. Secrecy is the ability to hide information from unauthorized individuals. Authenticity means being able to verify the source of information. Integrity means to validate that the message has not been tempered with in the stage of transmission[44]. All of these three criteria are achieved with public-key encryption which is an important part of computing today, frequently used in technological applications. It is an asymmetric type of encryption which means that instead of using the same key to encrypt and decrypt a specific message, like with symmetric encryption, the private key is used to encrypt and the public key is used to decrypt through a mathematical link between the private and the public key. Through the encryption secrecy is achieved. Anyone

intercepting the message that is not authorized (not in control of the private key), is not able (or again, it is infeasible) to read the message.

By sending a message digest with the full message (a compressed form of the message), the two parties can proof the integrity of the message[44]. Authenticity is achieved since only someone in control of the public key can relay a message which can be decrypted with the private key (and is not just mumble-jumble). This can be done simply by trying to decrypt the message with the public key and if something that makes sense comes out when decrypting, you are actually the holder of the private key and the person sending the message is in control of the corresponding public key.

4.3 Merkle Tree

A Merkle tree is a data-structure for providing integrity of files or data in general. Is used to store transactions in a block. Hash trees allow for the efficient and secure verification of contents of large data structures. A hash tree is build from hash values from data blocks. Starting from the bottom, we have four data blocks, each block gets hashed and we arrive at $H(A)$, $H(B)$, $H(C)$, $H(D)$.

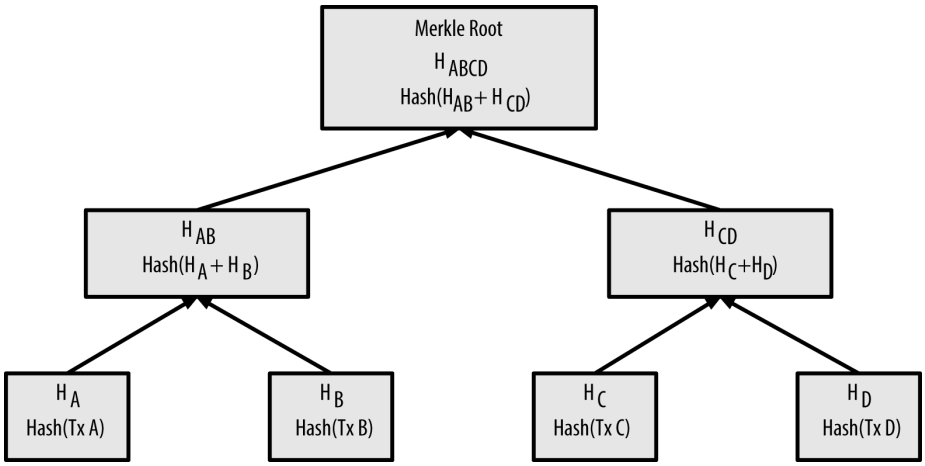

Figure 7: Merkle tree data structure.

Pairs of two of these hashes are then combined and passed through the hash function again. This process generates two separate and unique hashes which are each based on the combination of two hashes $H(AB)$ and $H(CD)$. These two hashes are again passed through the hash function and we arrive at the Merkle Root $H(ABCD)$[40].

4.4 Blockchain Features

When having a good understanding of the underlying cryptographic ideas, one can start to explore the specific features that together make up a Blockchain.

4.4.1 Blockchain

The Blockchain is the central object of the distributed peer to peer network. It is distributed among every client and is identical on every honest node[43]. How these Blockchains are kept synchronous is explained under Consensus Algorithms. The Blockchain contains many single blocks which are connected to each other. The first block in a Blockchain is called Genesis Block. The developer defines this block and fills it with arbitrary data. The Genesis block has to be the same for every node. A node without the correct genesis block will not be able to participate in the network. Each block has a height h and contains the hash of the previous block $h-1$.

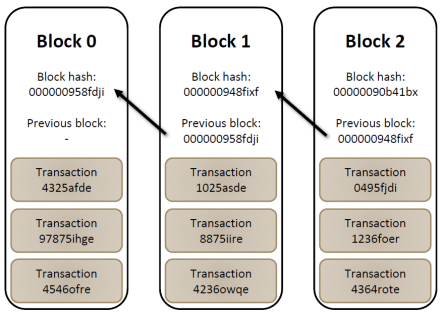

Figure 8: Blockchain data structure.

If any information is altered, the block header changes. Since the block header of the previous block is linked in the following block, the entire chain changes from that point forward and every following block will have to be recomputed. The same goes for the Merkle tree, if any data in the Merkle tree is changed, the header of the block will change[7]. This will reveal any compromise in the Blockchain. If the last hash of a node deviates from the rest, the node is not honest and will be excluded from the further process.

4.4.2 Block

A block is the data of the Blockchain which contains two objects: The Merkle tree of transactions and the block header. The header contains the crucial information about the block: The hash of the previous block, the version, a timestamp and the information about the search puzzle. The Merkle tree contains every transaction that is included by the miner up until the block-size limit. To build the Merkle tree, two transactions are concatenated and hashed. The results of each two concatenations are again connected and hashed, until there is only one hash left. The last hash is called the Merkle-root-hash and stored in the block-header.

4.4.3 Accounts and Transactions

An 'account' on the Blockchain is the combination of a public and a private key. The private key is the key to the account, whoever knows this 64 character hexadecimal number has full control over the given account[43]. A transaction is initiated when the person in control of the private key enters the public key of the person that they want to send something to and the amount to send. This action creates a transaction message, which is signed with the private key of the person that wants to send the funds. This transaction message includes the time, the amount, the public key of the receiver, the private key of the sender and other data. Using the public key of the sender, it can be verified that the person has indeed signed the transaction with the private key corresponding to the public

key and thereby is the account holder[43]. To include a transaction in a block, it is sent to all connecting nodes. The more nodes receive the transaction, the higher the probability that it is included in the successive block.

If someone wants to transfer a cryptocurrency to another participant, the individual has to be in control of unspent transactions. Transactions are either completely spent or unspent. It is impossible to spend only a fraction of a transaction. If this was possible, every miner would have to check the entire Blockchain for partially spent transactions. In order to make the process more efficient, everything has to be transferred. Funds that go above the given amount are then simply transferred back to the individual's account that sent the transaction, issuing a new unspent transaction. The output amount of every transaction is bound to a public signature. Only the individual with the matching private key to this public signature is allowed to spend the output. The sum of all inputs for a transaction has to be larger than or equal to the sum of all outputs. Otherwise, it would be possible to creation coins out of thin air. The overhead of the input and output is the transaction fee which is paid to the miner who proposes the block. It is possible to trace every transaction ever done one the Blockchain along with the initiator and receiver.

4.4.4 Process

To build a new block, the new transactions are validated according to the rules of the protocol, so that no false transaction gets stored in a block. Once this is finished the transaction is stored and built to a transaction tree along with meta information and other data. Once the block is built, the nodes participating in the network try to find the solution to a mathematical puzzle. How this puzzle works will be explained later. Once the solution has been found, the solution is included into the block. The node that found the solution adds the block to his chain and then announces the new chain with the latest block to the network. The other miners now validate the new block and check if all transactions are valid and execute them on their own copy of the network state. Thereby all nodes stay synchronous. Once this is finished, the process starts all over again.

An astute reader might now ask himself: What happens when two nodes find the solution at the same time? This is indeed possible and then the network has two valid chains of equal length with a different latest block since not all nodes get the same transactions at the same time. Both versions are send to the network and the nodes try to build atop the version that they received first, but also keep the other version in case it becomes larger[43]. As soon as one chain gets longer it is viewed as the correct chain and the rest of the network will again switch to this. Even those who were working on finding a solution for the other version will disregard their, now shorter chain, and regard the longer chain as the correct one. The block that is lost is considered an orphan block and the transactions from said block that are not already in the chain will return to the pool of unconfirmed transaction. This is also the reason why it is usually smart to wait for a few blocks to be completed after receiving a transaction before calling the transaction final. It is theoretically possible that two chains continue for a while when a block is found at the same exact time for multiple blocks subsequently. This chance is however insanely small and therefore negligible.

4.4.5 Consensus Algorithms in the Distributed Peer to Peer Network

The network is made of a distributed peer to peer network where everyone can partake. Every peer has the same privileges and is called a node. There are multiple consensus algorithms which allow for the nodes that make up the network to agree to a state of the network similar to a voting procedure. The algorithms each take a different approach. The underlying functionality that they all aim to achieve is that 51% of a network has to agree on a version of the ledger. This consensus plays a crucial role in the network. The main target is to achieve a joint state, meaning that everyone agrees to a certain state of the chain[6]. Since there is no central authority in the distributed system, in order to be able to decide which new blocks are valid and which are not every node has to decide individually whether it accepts a new block or rejects it. There are three basics of the consensus algorithm: The only valid chain is the chain that contains the correct genesis block, is made up of only valid blocks according to the rules of the network and is the longest chain[43]. The length of the chain is determined by the number of blocks. If a node has a longer chain, but the genesis block does is not correct or the blocks do not adhere to the protocol, the node is not honest. All other chains than the longest are rejected.

The theory of the Byzantine Generals dictates that at least 51% of the network have to be honest nodes for the distributed system to function properly[33]. The idea behind the consensus algorithm is that one node gets chosen randomly from the pool of all participating nodes. This node gets granted the right to fill the next block with transactions. However, since an individual could just enter with many different nodes, the process of increasing one's chances has to be very costly. How does this node actually get chosen? There are multiple consensus algorithms, which all aim to achieve the same thing. Following, the two most common will be elaborated.

Proof-of-work The proof-of-work algorithm involves scanning for a value that when hashed, such as with SHA-256, the hash begins with a number of zero bits. The average work required is exponential in the number of zero bits required and can be verified by executing a single hash. For the Bitcoin Blockchain, proof-of-work is implemented by incrementing a nonce in the block until a value is found that gives the block's hash the required zero bits. Once the CPU effort has been expended to make it satisfy the proof-of-work, the block cannot be changed without redoing the work. As later blocks are chained after it, the work to change the block would include redoing all the blocks after it[43]. It is important that the only way to find the solution to the puzzle is by brute-force. Thereby the only process of increasing one's chances to solve a puzzle is by investing in the equipment that is used to solve the puzzle. This tries ensures that no single entity can gain more hashing power than the rest of the network. Everyone always competes against the entire network. However, the problem with this is that it is a huge waste of energy since everyone is trying to solve the puzzle by doing completely useless computations by their processing units.

Proof-of-stake A different approach is the proof-of-stake algorithm. Here, control over parts of the currency makes it easier to propose new blocks. The

more stake someone has in the process, the higher are the individuals interest to keep the process running an honest. That is why the individual with the highest stake is chosen deterministically to propose the next block. This is of course only the case when punishing disingenuous behaviour. In order to take part in the proof of stake process, the individual has to freeze his balance. If the user is honest and does everything according to the protocol, the money is unfrozen and an extra reward is given. If it can be shown that the individual was dishonest, the stake is lost. The freezing of the stake can be seen as an opportunity cost as one is not able to spend it. The biggest advantage over proof-of-work is that there is no waste of resources[31]. However, it can be argued that for both algorithms, the rich will keep getting richer which will at some point lead to centralisation. In proof-of-work, those with the most money will buy more processing power and thereby increase their chances of solving a block. The same goes for proof-of-stake, where the stake just keeps getting bigger since the person with the highest stake will always get the reward. Neither of the two algorithms are perfect, but they are what we have to work with and for now, they work.

4.4.6 Trust-less System

The Blockchain is often characterised as a trust-less system. The argumentation behind this characterization is that the participants of the system do not need to trust other participants for it to work[43]. This is only true under the condition that the protocol has is flawless. No one has been able to find a flaw in the underlying Blockchain protocol, if someone had found a flaw, the systems would not exist any more. However, many Blockchain platforms allow not only the transfer of money but also the execution of so-called Smart Contracts[14]. Smart contracts are applications written by anyone that run on top of the Blockchain architecture. While the underlying architecture may be trust-less this does not automatically mean that the application on top of the architecture is also trust-less. The individual has to verify that the application can be trusted and does exactly what it should do.

4.4.7 Forks

The best example for an application where users had to much trust is the Decentralized Autonomous Organization or DAO. The DAO was a Smart Contract on the Ethereum Blockchain which functioned as a form of investor-directed venture capital fund. The idea was to have a decentralized business model for organizing both commercial and non-profit enterprises. The DAO had no conventional management structure or board of directors and all the logic was based purely on open-source code. The DAO was not tied to any particular nation-state and therefore stateless. This lead many regulators in confusion on how such a stateless fund would be dealt with[61]. The DAO was crowd-funded via a token sale in May 2016 and set the record for the largest crowd-funding campaign in history. However, in June 2016, users found an exploitable vulnerability in the DAO code which enabled them to siphon 1/3 of the DAO's funds to a subsidiary account. Through a security measurement in the code, the Ethereum community had 30 days to decide what they wanted to do. Two camps formed, one side was for leaving it as it is and losing thousand of investors money while

the other camp was for hard-forking the Blockchain. Hard-forking a Blockchain means to cut what came after a specific block in the chain and continue from that point on like it has never happened. Both sides had good arguments but in the end more than half of the network decided that the hard-fork was the right choice. This restored virtually all funds to the original contract. This was controversial, and led to a fork in Ethereum, where the original un-forked Blockchain was maintained as Ethereum Classic, thus breaking Ethereum into two separate active Blockchains, each with its own cryptocurrency. The main arguments of the camp that wanted to leave it as it is was that the intervention into the Blockchain showed that it was indeed not trust-less. The counter-argument for this is that the decision could only be made by more than half of the network agreeing to a certain decisions and that there was not on person or organisation that had the power to decide. However, this topic is very controversial in the Blockchain community and one could probably write another master thesis about it.

4.4.8 Private vs. Public Blockchain

When hearing Blockchain, most people think of Bitcoin or other public distributed ledger technologies. However, there are some Blockchains that are not as decentralized as the initial intention of the Blockchain intended them to be. These are called private Blockchains. They are controlled or overseen by a certain organisation or a consortium of organisations. Access to read and write is permissioned by one central authority. In my opinion, permissioned Blockchains offer nothing more than any traditional database can; any database can offer permissions, multiple input validation, multiple copies, append-only writes and logs of all people accessing it. In the following table, we get an overview of the differences in the two approaches.

	Private	**Public**
Access	Moderated	Anyone (read/write)
Identity	Known	Anonymous
Speed	Faster	Slower
Security	Moderated Access and known Identities	Proof of Work/ Proof of Stake/ etc.

Table 1: Private vs. Public Blockchains[67]

For the envisioned artefact, we want transparency. This already clashes with the idea behind private Blockchains. One could get the idea that private Blockchains are not really Blockchains and are merely an attempt of companies to add a buzzword for marketing purposes to their technology. However, this is up for debate.

4.5 Existing Blockchain Application Scenarios

In the following section, a few application scenarios of distributed ledger technologies are elaborated. The idea is to gain an understanding of what the possibilities of the Blockchain are.

Cryptocurrencies Cryptocurrencies are the most prominent use case of the Blockchain technology. Their basic idea is to allow for the unintermediated exchange of value online. However, cryptocurrencies are only one application of the Blockchain and with concepts like Ethereum, which provides a Turing complete scripting language that runs on the Blockchain, the possibilities for business scenarios seem endless[14]. Nevertheless, due to the novelty of this technology and the low number of experts, many organizations struggle to implement business applications that take advantage of the benefits that the Blockchain can provide.

Tokenization A token can be understood as a token at a festival. You buy the token with real money and can then use given token at a festival stand to buy food or a drink. The token thereby represents the value in money. On the Blockchain this happens frequently with the token representing a part of the company. So-called (Initial Coin Offerings) are the Blockchain way of a company going Public (Initial Public Offering)[15]. Tokenization describes the process of converting the rights to real-world assets into a digital token on the Blockchain representing said real-world asset.

Financial Services The most prominent use of the Blockchain after cryptocurrencies are financial services. Traditional systems are comparatively slow and error-prone and often require intermediaries to resolve conflicts and mediate the process, making them expensive. The Blockchain is seen by many as a transparent, cheap and more effective solution to provide financial services like Asset Management, Claims processing for Insurance and cross-border payments. One example is Asset Management is Omega One which is a platform that provides traders, investors and institutions with a decentralized and automated trade execution system that intelligently implements their trades across the world's crypto exchanges, shielding them from counterparty risk and significantly reducing the transaction cost of trades. Their vision is to become the worlds most advanced trading platform, providing institutions and serious individual traders with low-cost access to market tools usually reserved for the world's most technically-sophisticated hedge funds.

Commercial Applications A start-up that is making progress with Blockchain applications in the commercial world is Everledger/citeEverledger2018. Everledger focuses on the legitimation of objects. This is an application scenario where the Blockchain works well because of its immutable history and trust enabling consensus mechanisms. They offer their customers a distributed ledger of diamond transaction history and ownership verification. This is very useful for owners, law enforcement, insurance companies and claimants. The underlying idea of the system is to protect the customer against supply chain fraud and assists potential owners with the buying decision of a particular object. Leanne Kemp, the founder and CEO of Everledger explains that their ultimate goal is to track diamonds from mine to market, so that consumers are able to see if correct duties and taxes have been paid and make sure that a diamond is not a 'blood diamond', mined and traded in a war zone and has not contributed to human atrocity. The company plans to apply its technology to other big-ticket items, such as vintage cars, fine art and wine.

Developing Countries The potential of the Blockchain is also diverse in developing countries where the initial focus is on the trust element of the Blockchain. One company working on this is Factom[19]. They want to provide land registration for developing countries. This is one of the important application scenarios of the Blockchain in developing countries among other services like digital identity and finance for small-and medium sized enterprises. The problem in those countries is that often, the government and executive branches are fraudulent and bribing is a big issue. There is no possibility to register land and add a reliable title claim to your home. The Blockchain would allow people in places with poor registires, documents and rules of law to build trusted measures of their reputation. Giving the possibility to proof their identity by using their private key on the Blockchain. This would allow many of the world's two billion bankless individuals to permission banks to fulfil regulatory requirements and gain access to bank accounts, loans, and other financial services previously inaccessible to them. The potential of the Blockchain to revolutionize applications and drive global economic change is certainly there, but problems persist in wide-scale execution as the regulatory environment seems not yet completely ready.

Cryptokitties As silly as it may sound, but virtual collectable cats were the biggest thing on the Ethereum Blockchain for a while. CryptoKitties are virtual cats, comparable to trading cards, which can be breed with each other in order to create new collectable cats which can then be sold to other users[5]. The company Axiom Zen created the game on the Ethereum Blockchain, which accounted for over 10% of the network traffic during December 2017. The total sales volume of these tradeable virtual cats is $25.296.697 (to date) with the most expensive cat every being sold for $110.707 at the time of sale on the 7th of December in 2017[6]. The game is/was a first attempt at utilising the Ethereum Blockchain for leisurely and recreational purposes. In March 2017, Axiom Zen announced that CryptoKitties would be spun off into a company on its own, raising $12 million from multiple top angel investors and venture capitalist firms. The game is still under the 30 most gas intensive gas users on the Ethereum Blockchain today.

Consent Management There are already other attempts at utilising the Blockchain for Consent Management Systems. One of those solutions ins Dlock, Dlock is a platform which allows the person to regain the ownership of their personal data while making the companies GDPR-compliant. Dlock employs Blockchain for managing user consents making it impossible to challenge the fact that consent has been given or withdrawn. The user can manage their data via Dlock mobile app[7].

4.6 Blockchain Protocol Comparison

Since the release of Nakamoto's whitepaper and the Bitcoin Blockchain, many others have used the underlying idea to create protocols with different objectives

[5]https://www.cryptokitties.co/
[6]https://blockexplorer.com/news/cryptokitties-ethereum-blockchain-sell-100k/
[7]https://www.blockwise.org/our-solutions/

in mind. While Bitcoin was primarily designed for the use as disintermediated payment channel, others have taken the technological advances and combined them to new kinds of protocols that are able to do much more than just a cryptographic online payment channel. This chapter will look at the most prominent Blockchain protocols based on their market capitalization[16] and compare their different features in order to be able to determine which is the best protocol for this artefact. In order to determine which Blockchain architecture has the most to offer for the artefact, in the following the criteria that are derived from the literature about consent management and some ideas from the team at Bosch Software Innovations are elaborated shortly. The most important features of the underlying Blockchain architecture for this artefact are that it provides us with the freedom to develop a general-purpose platform, is decentralized, fast and transparent as well as reliable and has an active development community as well as active development on the architecture.

Functionality There are many different Blockchains which each aim to serve a different purpose. Since we want to develop an application on the platform we are looking for a general purpose platform that is not geared towards just fulfilling one specific task. We want to be able to write our own application on top of the architecture. Without this feature, the specific Blockchain is not suitable for the artefact.

Transparency The architecture should provide transparency for all parties, so that everyone can independently check whether the consent was given without having to rely on a third party. A Blockchain that only allows access for some is not suitable for the prototype.

Speed The speed of the underlying architecture is important since it is also a limiting factor for the speed of the application on top. However, by utilizing state-channels this problem can be solved so it is not the most important aspect. If the Blockchain is too slow and the Block creation time is too long, the process would slow down significantly which is unwanted.

Reliability The network has to be reliable with no down-times in order for our application to be available 24/7. If there are considerable down times of some sort this is a big minus. We are looking for a stable architecture that can provide a good underlying basis for our prototype.

Active Development The development of the platform should be very active to ensure that the project is moving forward and that it will not just be abandoned at any time soon. To quantify this number we will look at the activity in the GitHub repositories. More specifically, we look at the addition and deletion of code lines in the past months.

Dev. Community We want an active development community not only working on the development of the Blockchain itself but also developing on top of the Blockchain. Without an active development community the platform is not going to be around for long. To quantify this number we will look for the development communities on reddit and the amount of readers on this platform

as well as other community development platforms and their size might there not be an active reddit community for the platform.

	Bitcoin	Ethereum	EOS	Hyperledger Fabric	IOTA	NEO
Main application	Crypto Currency	Smart Contracts	Smart Contracts	Modular Framework	Machine to Machine	Smart Contracts
Smart contract functionality	limited	Turing complete	Turing Complete	Turing complete	no	Turing complete
Consensus mechanism	PoW	PoW (switch to PoS)	dPoS	PBFT	Tangle	dBFT
Transparency	yes	yes	yes	no	not entirely	yes
Reliability	high	high	medium	unknown	high	low
Transactions per second	7	20	3.000	3.500	1.000	1.000
Block time	10 minutes	15 seconds	0.5 seconds	/	/	30 seconds
Development community	non-existent	14.409	2.705	not public	2.484	1.221
Development activity Lines added/deleted	1.022 / 653	7.781 / 4.426	593 / 233	264 / 785	1.494 / 232	6.397 / 6.697

Figure 9: Comparison of different Blockchain architecture platforms.

Bitcoin is the most prominent Blockchain. This protocol is the mother of all Blockchains. But is it also the best for this prototype? The Bitcoin Blockchain is a public, permission-less Blockchain where anyone is allowed to join. The key features are as explained above, cryptographic hash functions, public-key cryptography, digital signatures, proof of work as consensus algorithm and the peer to peer network. With every node having the same — complete — information, the decentralized network allows for transactions without the need for a middleman[43].

The main functionality of the Bitcoin network is the cryptocurrency itself but it also allows for limited Smart Contract functionality. However, the scripting language is not Turing complete. The Bitcoin Blockchain is completely transparent and all code is open-source. Every block can be seen on the Bitcoin explorer[10]. The Bitcoin Blockchain is very reliable with no outage in since its creation. However, the Blockchain only allows for 7 transactions per second with a block time of 10 minutes which is very slow. The slow block time is a problem since it takes at least 10 minutes for a transaction to go through. There dose not seem to be an active development community for applications on the Bitcoin Blockchain. This is mostly the case since the Bitcoin Blockchain does only allow for limited Smart Contract functionality and is focused on the application as crypto-currency. The team behind Bitcoin is actively developing with 1022 lines added and 653 lines deleted from their GitHub project in the last month[9].

Ethereum By 2013, the community had thought up many different application scenarios for the Blockchain other than just the monetary use. However, people were building individual Blockchains for each and every application. Vitalik Buterin, then 19 years old, questioned the feasibility of that approach and came up with a general purpose Blockchain. He drastically changed the

approach from working like a Swiss army knife, where you have five different tools for five different categories of applications, to building a Blockchain that understands a general-purpose programming language. Ethereum is modelled after the smart-phone platforms iOS and Android, where everyone can write and publish an application that runs on the operating systems. The operating system on the Ethereum Blockchain is called the Ethereum Virtual Machine, it is a runtime environment for programs written in Solidity that run on the Ethereum Blockchain[14].

Programs build on the Ethereum Blockchain are called Smart Contracts, the term is misleading, since they are neither "smart" nor "contracts". Their basic functionality is determined by simple if-then statements. A basic example is a flight insurance product that pays out a policyholder when their flight is delayed for more than 2 hours. The policyholder would normally have to file a claim and then wait for the insurance company to process the request. With a Smart Contract, the user could file the claim on the Blockchain and the system would automatically determine whether the flight was delayed for more than 2 hours or not through a connection to a database which monitors flight times. If the claim is true, the policyholder will automatically be paid over the Blockchain.

Solidity is a Turing complete scripting language. The Ethereum Blockchain is public and therefore highly transparent. Every block can be inspected on the Ethereum Blockchain explorer online[62]. There has never been an outage since it is a decentralized system. Right now, the Ethereum Blockchain allows for up to 20 transactions per second. The network takes 15 seconds to build a new block. Ethereum has a highly active app development community with more than 14.000 users in the Ethereum development sub-reddit[8]. The Ethereum Foundation is also very active in further developing the Ethereum protocol with more than 7.000 lines of code added in the last month to the Ethereum core GitHub alone[17]. The Ethereum Foundation is working on many very promising independent projects simultaneously whose goal it is to further develop the protocol.

Hyperledger Fabric is a permission Blockchain network. The promise is to deliver truly modular, scalable and secure Blockchain solutions for industrial Blockchain solutions. It is a private consortium Blockchain. The consortium was formed by the Linux foundation. Today, more than 100 companies are cooperating, among those: IBM, SAP, Intel and Microsoft. The idea is to design and develop distributed ledger technologies for enterprise on the Blockchain basics. The custom made Blockchains are tailored to the needs of the customer. It is a modular framework for enterprise Blockchain solutions[9]. Since it is private, it is not transparent and we can not say anything about the speed, reliability and development community. Since Hyperledger is not a public Blockchain and you need to pay in order to use it, it is not an option for the project.

IOTA is a distributed ledger but it is not a Blockchain. However, since it is very popular it is still included in this comparison. The protocol focuses on machine to machine communication. Instead of using a Blockchain it uses a technology that the developers call the Tangle. This protocol is based off a

[8]https://www.reddit.com/r/ethdev/
[9]https://www.hyperledger.org/projects/fabric

mathematical concept known as directed acyclic graphs[32]. There is no need for miners and thereby also no transaction fees other than having to offer their computing power to validate two other transactions for one transaction. Everyone has an equal role in the network. When someone wants to make a transaction, the issuer must help to authenticate two previous random transactions. This allows for fast, inexpensive micro transactions that would be to expensive on other protocols. Since everyone that partakes in the network also has to verify two transactions, the speed increases as more users partake in the network making the transaction speed potentially almost unlimited[10]. However, this is not the case today since IOTA is very much in its infancy. With this infancy comes a weird feature of the network, the so called coordinator[11]. The code for the coordinator is not open-source which is a big no-go in the cryptocurrency/Blockchain community. The IOTA team has explained that they keep this part of the system behind closed doors to prevent copy-cats[12]. They claim that the coordinator will be removed as soon as the network is big enough. This coordinator helps to protect the network against certain attacks and the entire network is currently reliant on the coordinator. The coordinator also limits the network transaction speed to 1.000tps right now. The protocol does also not support Smart Contracts and is therefore not useful for the envisioned prototype.

EOS is a Blockchain and a Smart Contract platform. It is advertised as an operating system for decentralized applications just like the Ethereum Blockchain[13]. The platform uses delegated proof of stake (DPoS) as consensus mechanism. This consensus mechanism makes a trade-off between speed and decentralisation. In order to get faster and more scalable, it becomes more centralized (having 21 delegated block producers rather than an infinite number of miners in a proof of work model)[34]. There are very little resource on how to develop on the EOS platform and only a small development community with very little resources. The protocol is also very much in its infancy which also lead the Blockchain to experience a freeze a few days after its launch. Due to its infancy the reliability is only medium.

NEO is a Smart Contract based platform like EOS and Ethereum. It is described by the developers as a distributed network for the Smart Economy[14]. The pillars of the Smart Economy are digital assets, digital identity and Smart Contracts. Digital assets are programmable assets that exist in the form of electronic data. With Blockchain technology the digitization of assets can be decentralized, trustful, traceable and highly transparent as well as free of intermediaries. On the NEO Blockchain users are able to register, trade and circulate multiple types of assets. Proving the connection between digital and physical assets is possible through digital identity. Digital Identity refers to the identity information of individuals, organizations or other entities that exist in electronic form[50]. While this all sounds good at first, NEO has proved to be unreliable

[10] https://iota.readme.io/docs/whitepaper
[11] https://www.media.mit.edu/posts/iota-response/
[12]
[13] https://steemit.com/eos/@eosio/eos-io-technical-white-paper
[14] http://docs.neo.org/en-us/whitepaper.html

when the entire Blockchain went down after a single node disconnected temporarily on the 4th of March 2018. Currently all validator nodes are run by the NEO project since NEO does not use one of the usual consensus mechanisms but their own. The NEO architecture has a relatively small developer community with very little support due to the novelty of the platform.

4.7 Blockchain Conclusion

The Blockchain is a good platform for a consent management system. Here is why: The individual has been robbed of transparency when it comes to the consent process. In order for a consent platform to work it has to gain the user's trust. The Blockchain offers transparency, trustworthiness and security and is auditable by external parties. The shared reality aspect of the Blockchain makes the process see-through and accessible for everyone. Through the decentralized network there are no down-times and the system is always available and resilient against attacks since there is no single point of failure. When looking at the different Blockchain platforms, it becomes apparent that a private Blockchain is not an option since it simply does not provide the same qualities as a public Blockchain does. The only valid argument for private Blockchain in my opinion is the privacy aspect of data. On public Blockchains data is not private, however, this can be achieved by simply encrypting the data and only providing the access key to those that should have access. It might even be said that private Blockchains have no real place at all since they basically go against everything that the Blockchain stands for. Hyperledger is therefore not an option. IOTA is not an option since the protocol does not provide smart contract functionality. Even though other platforms like NEO and EOS sound very promising, they were/are in their infancy and can not deliver right now. Also due to the small amount of available developer resources. The Ethereum Blockchain provides the most extensive support system for developers and was therefore chosen for the purpose of this thesis. Not much research has been done on Consent Management Systems on the Blockchain. The work today is mostly theoretical. In the following section an artefact will be developed which aims to determine how such a prototype would look like and whether it makes sense to use the architecture.

5 Prototypical Application on the Ethereum Blockchain

The prototype aims to be a more up-to-date solution for consent management. Making the consent process more transparent and reliable and giving more privacy and control to the user. The consent management system was built on top of the knowledge gained in the literature research about consent management and in cooperation with the team at Bosch Software Innovations. The artefact aims to be a modern approach to consent management and thereby also explore the possibilities of business applications for the Blockchain. As explained in the previous section, the Blockchain has some unique advantages compared to traditional systems. With this artefact, the opportunities of the Blockchain are explored in order to be able to determine whether such an approach is feasible. In the following section the different iterations of the prototype and the most important functionality of each iteration will be explained with snippets of the code. The code shown in each iteration is not complete and some lines are abstracted for brevity.

The rest of the chapters is blocked by request of Bosch until the 1st of July 2019.

6 Conclusion and Outlook

The goal of this master thesis was to answer the following question: How can user consent be managed in a transparent and straightforward system, utilising the Ethereum Blockchain, where the user has control over what happens to their data beyond organisational boundaries? The question was split into four sub-questions, these questions will be answered in the following section:

6.1 Conclusion

SQ1 & SQ2: What does Consent Management look like today and what is the problem with Consent Management? The most common consent mechanisms are Terms and Conditions, End User Licence Agreements (EULAs) and Terms of Service (ToS). When agreeing to the End User License Agreement, the individual usually only has to click the "I Accept" button. This interaction represents the moment of consent in which the user is indicating that he/she is consenting to whatever is in the EULA, ToS or T&C[37]. Research shows that less than 1% actually pause to read what's written in these agreements and that those who do, do not spend enough time to be able to have digested even a fraction of the agreement[5]. The contemporary articulation of consent has been stretched thin to the point of breaking[37]. Consent is not clear, it's often full of illegible terms and conditions full of legalese. Users do not know or understand what they are actually giving consent to. EULAs, the main form of consent today, therefore seem to be completely useless in informing the consumer, only as protection for liability for the companies. Consent agreements have to become easier to read, assess and compare in order to achieve the goal of making consent more clear[5]. Other, more modern approaches to consent management are systems like BMW's CarData or Mercedes consent system. They are however part of the data provider's architecture. Such data providers are in charge of both, the personal user data is and the individual consent agreements corresponding to the data. The resulting centralisation of responsibilities increases the need for trust in the data provider. In addition, it is not possible for third parties to access and validate an individual user consent.

Existing Consent Management Platforms are focused on one scenario: Online data collection and sale to advertising firms. Another important issue with consent management is that it ignores severe human cognitive problems that impair the ability of the individual to make rational and informed decisions about the benefits and costs of disclosing their personal data [58]. To overcome these human cognitive problems, privacy notices have to become more clear-cut and the individual, as well as companies, have to become more aware of the personal data that is traded day by day, its value and the security risks. During the sign up for such a service, a check for age and mental illness has to take place in order to protect the individual against exploitation.

The data has to be presented in a very straight-forward way without too much distraction and only through an affirmative action should consent be valid. The goal should be to have a good transaction framework with more direct information disclosure of accurate and relevant information, rather than a general full disclosure that could easily flood the user with too much information, resulting in a confused or ignorant decision rather than an informed one[21]. It

is also crucial that consent data is transparent and that there is no middleman who controls the process of consent as well as the data and can allow or disallow access. The consent data should be visible for the entire consent process chain, so that everyone can check, individually, whether consent was given. This expression of consent has to happen in an explicit way, where the individual is not coerced into agreeing to something because they do not see that a box is already checked. The action to give consent has to be affirmative and unambiguous. It should also be possible to get an overview of the consent history and to revoke consent as easily as it is to give consent.

SQ3 & SQ4: What is the Blockchain and wow can be used for Consent Management System? The Blockchain is an append-only log, storing transactions. All data is fully replicated across a large number of peers, called nodes. Data is combined in immutable blocks which are deterministically verifiable using the Blockchain data structure. The Blockchain is fully decentralized and does not rely on a third party for trust. Immutability is achieved using hashing. The data is replicated across the entire network of peers, leaving everyone with the same information. Consensus is reached through a Byzantine proof algorithm like proof of work (pow) or proof of stake (pos). Every node participating in the network verifies every transaction. The integrity and anonymity of the network is achieved through the clever use of cryptography. The Blockchain is a good platform for a consent management system. Here is why: The individual has been robbed of transparency when it comes to the consent process. In order for a consent platform to work it has to gain the user's trust. The Blockchain offers transparency, trustworthiness and security and is auditable by external parties. The shared reality aspect of the Blockchain makes the process see-through and accessible for everyone. Through the decentralized network there are no down-times and the system is always available and resilient against attacks since there is no single point of failure.

When looking at the different Blockchain platforms, it becomes apparent that a private Blockchain is not an option since it simply does not provide the same qualities as a public Blockchain does. The only valid argument for private Blockchain, in my opinion, is the privacy aspect of data. On public Blockchains data is not private, however, this can be achieved by simply encrypting the data and only providing the access key to those that should have access. It might even be said that private Blockchains have no real place at all since they basically go against everything that the Blockchain stands for. Hyperledger is therefore not an option. IOTA is not an option since the protocol does not provide Smart Contract functionality. Even though other platforms like NEO and EOS sound very promising, they were/are in their infancy and can not deliver right now. Also due to the small number of available developer resources. The Ethereum Blockchain provides the most extensive support system for developers.

6.2 Discussion & Future Research

This master thesis has shown that it is possible to use the Blockchain for consent management and therefore for many business applications. What works well is the transparency and open platform that the underlying architecture of the Blockchain is able to provide. One drawback that complicates the development

of such an application is however exactly this transparency. The developers have to think about the drawbacks of putting everything out in the open and have to come up with solutions to the challenge of keeping sensitive data private.

One thing that proved to be very difficult is measuring time. There is no global clock build into the Blockchain which allows to determine a certain time frame. When wanting to give permission for a certain period, one has to rely on the Blocktime which is not a precise unit at all. Therefore a function had to be added which lets the administrator change the block time. This however decreases the "trust-less" property of the architecture in that it allows the administrator to tamper with the duration of given consents.

Another challenge of the Blockchain architecture and its use for business applications are the factors of efficiency and the costs. Utilising a Blockchain is not about efficiency: Every transaction is not only processed by one processing unit but by an entire network. One operation is therefore done hundreds if not thousands of times. It should be obvious that this is expensive. Practitioners have to be aware of this and really think if the Blockchain is able to add so much value to their envisioned application, that this inefficiency is justified.

There are many technical and theoretical challenges that have to be tackled in the future for the Blockchain to become production ready. The cost for a Blockchain application is much higher compared to a traditional system due to the inefficiency of multiple processing units processing the same transaction. Another problem is the speed of the network. The consensus mechanisms in the protocol today either go towards centralization or become slower. There are solutions on the horizon which solve these problems, however, their implementation lies in the future.

Another issue is the market introduction for such an application. Which will most likely prove to be very challenging, not only because of the resistance from existing companies but also from a regulatory stand-point. The legality of Smart Contracts is a topic that has yet to be fully explored. All of these topics are very interesting and each and every one requires more work in the future.

6.3 Research Limitations

The limitations of this master thesis were mainly that the Blockchain is very much in its infancy. This complicated many aspects of the research as well as implementation process. Many of those problems will probably be solved in the future but for now they still exist. With the infancy of the Blockchain, another limitation was that practitioners are far ahead of the research community. This complicated the search for credible sources and less credible conference proceedings and white papers had to be used. Another limitation was my limited knowledge of programming and information systems and the limited available information on programming in Solidity. Coming from a business background, everything I know about programming and IT is self-taught which proved to be challenging at times. Not having the background knowledge and being informed about the standard approaches was very time intensive.

6.4 Recommendations for Practice

A company thinking about utilising the architecture has to ask themselves two important questions: What value does the Blockchain add? & Is the technology

mature enough for our envisioned application?. Organisations often get drawn into hype-topics since they are good for marketing. A prime example for this are private Blockchains which stray so far from the actual idea of the Blockchain that calling these protocols Blockchain is almost offensive.

My recommendation for practitioners is to wait and see how the technology develops in order to not waste money on the development of such a system which might in the end not be able to function due to one of the many challenges elaborated above. The Blockchain technology is very much in its infancy. There are many promising ideas that solve almost all of these problems, however, until those are resolved it is simply too early to dedicate time and resources to such a project. While a consent management system on the Ethereum Blockchain is possible with the development of a Smart Contract that has the underlying logic to handle these processes, it stands to argue whether now is the right time to spend time and resources on the development of such a platform.

References

[1] Alessandro Acquisti and Jens Grossklags. Privacy and rationality. In *Privacy and Technologies of Identity*, pages 15–29. Springer, 2006.

[2] Alessandro Acquisti, Curtis R Taylor, and Liad Wagman. The Economics of Privacy. *Ssrn*, 54:442–492, 2015.

[3] Annie I Anton, Julia B Earp, Qingfeng He, William Stufflebeam, Davide Bolchini, and Carlos Jensen. Financial Privacy Policies and the Need for Standardization. *IEEE Security and Privacy*, 2(2):36–45, 2004.

[4] Article 29 Data Protection Working Party. Guidelines on Consent under Regulation 2016/679. *October*, (April):1–11, 2003.

[5] Yannis Bakos, Florencia Marotta-Wurgler, and David R Trossen. Does Anyone Read the Fine Print? Consumer Attention to Standard-Form Contracts. *The Journal of Legal Studies*, 43(1):1–35, 2014.

[6] Arati Baliga. Understanding blockchain consensus models. *Persistent*, 2017.

[7] Roman Beck, Jacob Stenum Czepluch, Nikolaj Lollike, and Simon Malone. Blockchain – the Gateway To Trust- Free Cryptographic Transactions. *Twenty-Fourth European Conference on Information Systems (ECIS)*, pages 5–16, 2016.

[8] Steve Benford, Matt Adams, Ju Row Farr, Nick Tandavanitj, Kirsty Jennings, Chris Greenhalgh, Bob Anderson, Rachel Jacobs, Mike Golembewski, Marina Jirotka, Bernd Carsten Stahl, Job Timmermans, and Gabriella Giannachi. The Ethical Implications of HCI's Turn to the Cultural. *ACM Transactions on Computer-Human Interaction*, 22(5):1–37, 2015.

[9] Bitcoin. Bitcoin GitHub Repository Monthly Statistic, 2018.

[10] Bitcoin. The Blockchain Explorer, 2018.

[11] BMW. No Title, 2018.

[12] Christian J Bonnici and Lizzie Coles-Kemp. Principled electronic consent management: A preliminary research framework. *Proceedings - EST 2010 - 2010 International Conference on Emerging Security Technologies, ROBOSEC 2010 - Robots and Security, LAB-RS 2010 - Learning and Adaptive Behavior in Robotic Systems*, pages 119–123, 2010.

[13] David Budgen and Pearl Brereton. Performing systematic literature reviews in software engineering. *Int. Conf. Soft. Engin.*, page 1051, 2006.

[14] Vitalik Buterin. A next-generation smart contract and decentralized application platform. *Etherum*, (January):1–36, 2014.

[15] Christian Catalini and Joshua S Gans. Initial coin offerings and the value of crypto tokens. Technical report, 2018.

[16] Coin Market Cap Team. Coin Market Capitalization, 2018.

[17] Ethereum Foundation. Ethereum GitHub Repository Monthly Statistic, 2018.

[18] European Parliament and Council of the European Union. Regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (Data Protection Directive), 2016.

[19] Factom. Making the world's system honest, 2018.

[20] Ruth R Faden and Tom L Beauchamp. *A history and theory of informed consent*. Oxford University Press, 1986.

[21] Catherine Flick. 8 . Informed consent in information technology : Improving end user licence agreements. (May 2017):126–154, 2013.

[22] B Friedman, D C Howe, and E Felten. Informed consent in the Mozilla browser: Implementing value-sensitive design. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2002-Janua(c):1–10, 2002.

[23] Batya Friedman, Edward Felten, and Lynette I Millett. Informed Consent Online. *DisClosure*, pages 1–8, 2000.

[24] Batya Friedman, Peyina Lin, and Jk Miller. Informed consent by design. *Security and Usability*, (2001):503–530, 2005.

[25] Alan Hevner and Samir Chatterjee. *Design Science Research in Information Systems*, volume 22. 2010.

[26] Bjørn Hofmann. Broadening consent—and diluting ethics? *Journal of Medical Ethics*, 35(2):125–129, 2009.

[27] Eric Hughes. A cypherpunk's manifesto. *URL (accessed 21 March 2018): http://www. activism. net/cypherpunk/manifesto. html*, 1993.

[28] IAB Tech Lab. GDPR Transparency and Consent Framework, 2018.

[29] John P A Ioannidis. Informed consent, big data, and the oxymoron of research that is not research. *The American Journal of Bioethics*, 13(4):40–42, 2013.

[30] Ghassan O Karame, Elli Androulaki, and Srdjan Capkun. Double-spending fast payments in bitcoin. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 906–917. ACM, 2012.

[31] Sunny King and Scott Nadal. PPCoin: peer-to-peer crypto-currency with proof-of-stake (2012). *URL https://peercoin. net/assets/paper/peercoin-paper. pdf.[Online*, 2012.

[32] B Kusmierz. The first glance at the simulation of the Tangle: discrete model, 2017.

[33] Leslie Lamport, Robert Shostak, and Marshall Pease. The Byzantine Generals Problem. *ACM Transactions on Programming Languages and Systems*, 4(3):382–401, 1982.

[34] Daniel Larimer and Brendan Blumer. EOS White Paper. 2018.

[35] Ewa Luger. Consent reconsidered; reframing consent for ubiquitous computing systems. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*, page 564, 2012.

[36] Ewa Luger, Stuart Moran, and Tom Rodden. Consent for All: Revealing the Hidden Complexity of Terms and Conditions. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, page 2687, 2013.

[37] Ewa Luger and Tom Rodden. An informed view on consent for UbiComp. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 529–538. ACM, 2013.

[38] Ewa Luger and Tom Rodden. The Value of Consent. pages 1–25, 2014.

[39] Yvonne McDermott. Conceptualising the right to data protection in an era of Big Data. *Big Data & Society*, 4(1):205395171668699, 2017.

[40] Ralph C Merkle. Protocols for public key cryptosystems. In *Security and Privacy, 1980 IEEE Symposium on*, page 122. IEEE, 1980.

[41] M C Mont, Siani Pearson, Gina Kounga, Yun Shen, and Pete Bramhall. On the Management of Consent and Revocation in Enterprises : Setting the Context. 2009.

[42] Alistair Morrison, Donald McMillan, and Matthew Chalmers. Improving consent in large scale mobile HCI through personalised representations of data. *Proceedings of the 8th Nordic Conference on Human-Computer Interaction Fun, Fast, Foundational - NordiCHI '14*, pages 471–480, 2014.

[43] Satoshi Nakamoto. Bitcoin: A Peer-to-Peer Electronic Cash System. page 9, 2008.

[44] James Nechvatal. Public-key cryptography. (April), 1991.

[45] Oxford Dictionaries. Definition of Consent, 2018.

[46] Aditya Pakalapati. A Flexible Consent Management System for Master Person Indices. 2012.

[47] Ken Peffers, Tuure Tuunanen, Marcus A Rothenberger, and Samir Chatterjee. A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3):45–77, 2007.

[48] Eugenia Politou, Efthimios Alepis, and Constantinos Patsakis. Forgetting personal data and revoking consent under the GDPR: Challenges and Proposed Solutions. *Journal of Cybersecurity*, (April):1–20, 2018.

[49] Paul Prinsloo and Sharon Slade. Student privacy self-management. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK '15*, pages 83–92, 2015.

[50] N E O Project. NEO White Paper, 2018.

[51] Byron Reeves and Clifford Ivar Nass. *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge university press, 1996.

[52] H W J Rittel and M M Webber. Planning problems are wicked problems. N. Cross (Ed.). Developments in Design Methodology (pp. 135-144), 1984.

[53] P Rogaway and T Shrimpton. Cryptographic hash-function basics: Definitions, implications, and separations for preimage resistance, second-preimage resistance, and collision resistance. In *Fast Software Encryption*, volume 3017, 2009.

[54] John Rose, Olaf Rehse, and Björn Röber. The Value of Our Digital Identity, 2012.

[55] Mark A Rothstein and Abigail B Shoben. An unbiased response to the open peer commentaries on "does consent bias research?". *The American Journal of Bioethics*, 13(4):W1–W4, 2013.

[56] Andy Siddaway. What is a systematic literature review and how do I do one? *University of Stirling*, (Ii):1–13, 2014.

[57] Herbert A Simon. *The sciences of the artificial.* MIT press, 1996.

[58] Daniel J Solove. Privacy Self-Management and the Consent Paradox. 1880, 2012.

[59] Fiona Stevenson, Nigel Lloyd, Louise Harrington, and Paul Wallace. Use of electronic patient records for research: views of patients and staff in general practice. *Family practice*, 30(2):227–232, 2012.

[60] Katherine J Strandburg and Daniela Stan Raicu. *Privacy and Technologies of Identity: A cross-disciplinary conversation.* Springer Science & Business Media, 2005.

[61] Swinburne University of Technology. The radical DAO experiment, 2016.

[62] Team Etherscaners. The Ethereum Block Explorer, 2018.

[63] Wojciech A Trybulec. Pigeon Hole Principle. 2(1):6–10, 1990.

[64] Viswanath Venkatesh, James Y L Thong, and Xin Xu. Consumer Acceptance and Use of Information Technology : Extending the Unified Theory. *MIS Quarterly*, 36(1):157–178, 2012.

[65] Edgar A Whitley. Informational privacy , consent and the " control " of personal data. *Information Security Technical Report*, 14(3):154–159, 2009.

[66] Edgar A Whitley. Informational privacy, consent and the "control" of personal data. *Information security technical report*, 14(3):154–159, 2009.

[67] Gavin Wood. Blockchains: What and Why, 2016.

[68] World Economic Forum. *Personal data : The emergence of a new asset class*. 2011.

7 UTAUT Questionnaire