The impact of control, trust and risk on business participation in data marketplaces

Floris Kool

MSc Thesis
Complex Systems Engineering and Management
May 2023

The impact of control, trust and risk on business participation in data marketplaces

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Complex Systems Engineering and Management**

Faculty of Technology, Policy and Management

by

Floris Kool

Student number: 4975243

To be defended in public on May 15th 2023

Graduation committee

Chairperson	: Dr. ir, G.A. de Reuver, Section ICT
First Supervisor	: Dr. ir, G.A. de Reuver, Section ICT
Second Supervisor	: Dr. ir, Z. Roosenboom-Kwee, Section ETI

Acknowledgments

I would like to thank my graduation supervisor Mark de Reuver for all the valuable feedback and support over the last months. It turned out that I did not enjoy doing academic research much and I struggled much more than I expected. The nearly weekly feedback meetings really made the difference.

Additionally I want to say a big thank you to Wirawan Agahari, who helped me out in the formation of the theory and offered the leftover budget of his previous study. Without this I surely would not have been able to collect such good responses. I'm also very grateful for the help of my second supervisor Zenlin Roosenboom-Kwee and all her feedback.

Lastly I want to thank my girlfriend Nienke and my family for all their support.

Executive summary

Business data sharing can generate many benefits for the manufacturing industry. Sharing data is becoming easier, as data marketplaces allow for data sharing with more companies. Unfortunately the adoption of these marketplaces and the amount of data shared is lacking. There are several factors that play into this such as: the benefits of sharing data, the amount of data available, the interoperability of IT systems and trust between users. Data marketplaces work differently from other information sharing platforms (such as supply chain platforms), because they are more open. This openness allows data to be shared with many more companies, bringing more potential benefits, but also increasing the risk. On data marketplaces there typically is less trust between users, however this lack of risk can be compensated by adding more control mechanisms. For these reasons I want to address the research question:

How does the interplay between control, trust and risk affect a company's decision to share data in a data marketplace?

I will answer this question by performing a literature review and a quantitative study. The literature review resulted in a conceptual model that describes the relationships between 6 constructs: perceived Control, trust in other users, trust in platform, sensitivity of data, perceived risk and willingness to share data. The perceived control is about control from the perspective of a data provider. It can be influenced by adding in more control mechanisms. Adding more control should give a data provider less risk when sharing data. Trust in other users is also proportional to the risk of sharing data. In data marketplaces I expect there is less trust with other users, which needs to be compensated with more control. Trust in the platform should help reduce risk of sharing data directly and is also contributing to the perceived control. If a person has low trust in a platform, they are unlikely to perceive a lot of control over their data. Sensitivity of data should also have an impact on the risk of sharing data. If the data is not very sensitive, it's unlikely the data provider will experience a lot of risk. In the conceptual model perceived risk is the only factor directly influencing the willingness to share data.

To test the conceptual model I conducted an online questionnaire for management in the manufacturing industry. The respondents were given two scenarios in which they had a lot of industrial internet of things (iIoT) data available. In one scenario they would experience more trust and more control, and the other would be a more open platform. At both scenarios I measured their stance on the 6 constructs from the conceptual model. The response collection happened from Friday 10th of March 2023 until Sunday 12 of march 2023. After filtering I had 295 complete responses. After some minor modifications to the measurement model, I used structural equation modelling on the data from the questionnaire to test the relationships defined in the conceptual model. 5 out of the 6 relationships were confirmed by my data, with only the relationship between trust in the platform and perceived risk being insignificant.

Answering the research question, this study has shown that increasing trust in other users and perceived control can really help decrease the risk of sharing data, which in turn is an important predictor of willingness to share data. However, the sensitivity of data can also have an influence over risk. This insight is new in the field of data marketplaces. We now have a more clear picture of how control and trust can be used to lower the risk and in turn increase the willingness to share data. This study also provided some data that confirms these relationships in the context of the manufacturing industry. This data shows that companies are willing to share their data, if preconditions such as a fair price are met. These results contribute to the overall literature on willingness to share data in a marketplace and on the literature on control mechanisms, trust and risk. The insights from this study can be useful for people designing data marketplaces and regulators that want to encourage the sharing of data.

Contents

List of Figures

List of Tables

1 Introduction

Business are generating more and more data. An example of this is the manufacturing industry transitioning to industry 4.0. A concept based on 4 emerging technologies: Internet of Things, Cloud services, Big Data and Analytics (Lasi et al., 2014; Frank et al., 2019). This available data brings many opportunities to improve the quality of their products, reduce test and calibration time, reduce costs and perform predictive maintenance (Lade et al., 2017). If companies were to share their data, the benefits would increase much further (Dai et al., 2020). Unfortunately usage of the data and especially data analytics is still lacking (Frank et al., 2019). The willingness to share data is a well studied, but complicated matter. Companies would only ever share data if the benefits are clear to them (Li et al., 2006), they trust the partner company (Chen et al., 2014) companies are ready to share information in terms of organization and ICT (Azarm-Daigle et al., 2015).

Digital platforms facilitate new ways to share data between companies, with new benefits (Beverungen et al., 2022). Examples of this are data marketplaces where businesses can buy and sell data. It is still very unclear how willing companies are to participate in such marketplaces. Literature shows there are a lot of barriers for sharing data between companies (Kuan & Chau, 2001). It will be interesting to see how digital platforms change the way companies approach sharing data. I specifically expect that the perceived risk of sharing data will change with the rise of digital platforms, as companies might share data to companies they don't trust as much. However, digital platforms are able give new control mechanisms that could mitigate this lack of trust (Reimsbach-Kounatze, 2021). This research aims to find out why companies would share data in a marketplace and what barriers there are. Specifically focusing on the relationship between control, trust and perceived risk when sharing data through an open platform. I will do this by performing a survey with the manufacturing industry as a target group and analyzing the results using structural equation modeling.

This chapter defines what will be studied in this master thesis. Starting off I will describe scientific contribution of the study, by describing the knowledge gap. Then I will give a clear research objective, followed by the main research question. After this I will describe how this question will be answered, by describing the research approach and providing sub questions. This is followed by an outline of the thesis. At the end of this chapter a description will be provided that explains why this is a suitable topic for a Complex Systems Engineering and Management master thesis.

1.1 Knowledge gap

Previously data sharing through platforms was done between (trusted) business partners. This has been extensively researched. Data marketplaces work differently compared to these platforms, by connecting many unknown parties. This can create new value propositions, but can also create new risks. Making the choice to share data on an open marketplace is different from the choice to share data with a trusted business partner. This choice has not yet received a lot of attention from researchers. This is expressed by Abbas et al. (2021). Who performed a literature review on data marketplaces, identifying that there is not a clear understanding why the commercialization of data marketplaces is unsuccessful and recommend further research on the service aspects of this. De Prieelle et al. (2022) did a study on Internet-Of-Things (IoT) data sharing, identifying a number of factors influencing the adoption of data marketplaces and their relative importance. Kazantsev et al. (2022) did an exploratory study to find barriers for information sharing between SME's. They looked at companies within the European aerospace industry and found 5 groups of barriers: impeding: market transparency, access to orders, partner trust, contracting and data sharing and coordination. The study was however performed on a limited scale with a specific low-volume high-variability sector within the manufacturing industry. Holler et al. (2019) performed an exploratory study on manufacturing companies willingness to share data, but so far has only published a Research-in-Progress paper. Their preliminary results show a wide range of factors: Trust and established relationship, type and frequency of product data and data security practices. Overall there has not been a lot of research on the adoption of data marketplaces, in the manufacturing industry and in general. This study aims to contribute to this research, by specifically focusing on the impact of risk on the adoption of data marketplaces.

The perceived risk of trading data in a marketplace is very different when compared to trading it with business partners. There is likely less trust between data provider and data consumer, which traditionally would make data sharing unlikely (Müller et al., 2020). However these platforms can provide tools that give control back to the data provider (Reimsbach-Kounatze, 2021). This interplay between trust and control impact the way a company perceives the risk of exposing competitive advantage or breaching privacy by sharing personal data. Agahari et al. (2022) looked into these factors in the context of the automotive industry with a trustless technology called Multi-Party Computation. They found that trust in other parties and risk become less important when this technology is used, but trust in technology and control mechanisms become more important. These studies show there is a relation between trust, control and risk when sharing data. It could use more attention as it's not been studied in the context of a more open data marketplace.

Out of all the adoption factors of data marketplaces I decide to focus on control, trust and risk. This is because I believe the relationship between these factors and the impact they have is very different in the context of a data marketplace compared to sharing data with known business partner. So far research on adoption factors of data marketplaces has revealed similar results when compared to adoption of traditional data sharing platforms (with trusted business partners). However it would seem that sharing in a data marketplace would bring many other risks with it. The trust and relationship between companies would be very different and the control mechanisms should have a larger role in mitigating the risk. This is why I specifically want to find out how these factors influence each other and how they influence the willingness to share data.

Before diving into the relationship between these factors, it's important to look at them separately. The perceived risk and perceived control might be different from, actual levels of risk and control. So far there is no standard way of increasing control or trust, or decreasing risk. I need to research into what mechanisms are available to increase control, trust or decrease risk and evaluate their effectiveness and usability.

As described, there are some theories developed on why companies are participating in data marketplaces. Using these theories there has been some exploratory qualitative work done with in depth interviews and case studies (Holler et al., 2019; Nazifa and Ramachandran, 2019; Agahari et al., 2022). What's lacking is validation that this really is what influences companies decisions. A qualitative approach, as describe by Creswell (2009) can help validate these results. The need for a quantitative approach is also expressed in the extensive literature review by Abbas et al. (2021). A quantitative study can also provide better understanding how different factors influence a companies decision. By zooming in on the relationship between control, trust and risk, we will get a better understanding how they contribute to a company's decision to share data.

To summarize: There needs to be a better understanding of why companies are willing to share data, specifically the control trust and risk require more attention. Mostly because they differ a lot in the recent open marketplaces, compared to traditional platforms that allow data sharing with trusted business partners. A quantitative research can give insights into how control, trust and risk influence a companies decision to share data.

1.2 Objective

This research aims to define and validate relationships between control, trust and risk when sharing data in a data marketplace and how they impact the willingness to share data. This includes: figuring out what mechanisms can affect the perceived control, trust and risk, then defining the relationships between the factors, and finally I aim to validate these relationships through the use of a questionnaire in the manufacturing industry and structural equation modelling to analyze the results. I will look at the relationships from the perspective of data providers, but this information can be used in the design of data marketplaces.

1.3 Research question

To create a better understanding of why companies participate in data marketplaces, validate the result of the exploratory studies on this subject and get further insight on the relationship between control, trust and risk, I will do a quantitative study to answer the question:

How does the interplay between control, trust and risk affect a company's decision to share data in a data marketplace?

1.4 Research approach

Before doing a quantitative study I want to get a clear picture of the theory. To start this I will make an overview of all the important factors that influence data marketplace adoption. After that I will zoom in on the relationship between control, trust and risk, also looking into how this is different for open platforms compared to traditional business to business platforms. Special focus will be given to different control mechanisms. Once the theory is fully developed I will test this by surveying companies from various sectors within the manufacturing industry. I expect that these results will give more insights on why companies share data. The results will also likely show differences between companies in different sectors. Following this process I will answer the following sub questions:

1. *What mechanisms are available for managing control and trust?*
2. *How do control and trust impact a company's perceived risk of sharing data?*
3. *How do different control mechanisms influence the perceived risk of sharing data?*
4. *How does the trade off between control mechanisms and trust work?*
5. *How does the perceived risks impact a company's decisions to share data?*
6. *How does the sensitivity of data and contextual factors influence the relationship between control, trust, risk and willingness to share data?*

1.5 Research structure

Answering the sub questions will be done in 5 steps. First a literature search will be performed to find out what drives adoption of data marketplaces and what the relations between control, trust and risk are. Secondly a structural model will be created along with hypotheses that need to be tested using a survey. After that a survey will be designed that aims to validate the theoretical model. The results of this survey will be analyzed using structural equeation modelling and interpreted. appendix A shows a complete research flow diagram describing this process. Figure 1 shows a simplified version of the research structure. The roman numerals at the bottom of each step in the orange boxes is a reference to the phase(s) in the research flow diagram (appendix A). The Cx numbers in the blue boxes refer to the thesis chapter(s) that describe this step.

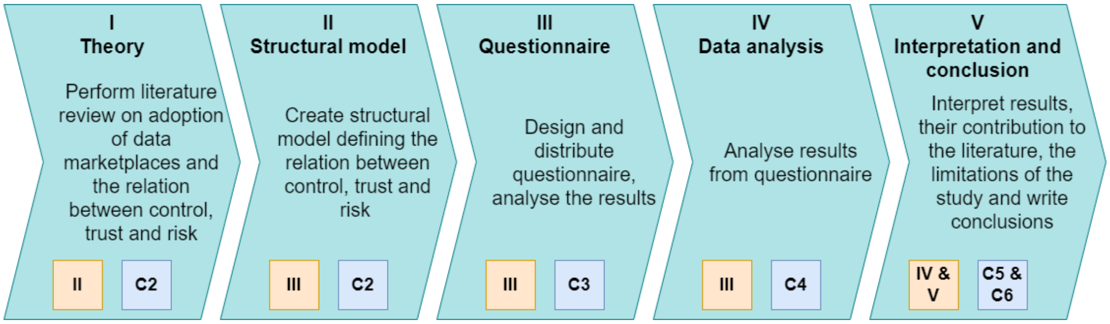

I
Theory
Perform literature review on adoption of data marketplaces and the relation between control, trust and risk
II
C2
II
Structural model
Create structural model defining the relation between control, trust and risk
III
C2
III
Questionnaire
Design and distribute questionnaire, analyse the results
III
C3
IV
Data analysis
Analyse results from questionnaire
III
C4
V
Interpretation and conclusion
Interpret results, their contribution to the literature, the limitations of the study and write conclusions
IV & V
C5 & C6

Figure 1: Research structure

1.6 Link to Complex Systems Engineering and Management

This research topic is suitable for a Complex Systems Engineering and Management master thesis, because data ecosystems are complex sociotechnical systems with technical components and many institutions regulating the conflicting needs of many different actors. Taking a scientific approach the participation of companies will be analyzed using multiple perspectives from various disciplines to arrive to balanced conclusions. While this study looks at data marketplaces from the perspective of data providers, the design of the data platform will heavily influence the outcomes of the study. Defining different design options will therefor be a large part of the investigation. The outcomes of the study should create better understanding of why data owners will share data, and also how platform design can contribute to this.

With this chapter I've sketched what issues in the adoption of data marketplaces are, why control, trust and risk have influence over this and why this further research into this is important. The chapter also outlines how this will be done by providing sub questions, an approach and a research structure. The next chapter will go deeper into the theory. Looking into what the relationships between control, trust and risks are, with the aim of creating a conceptual model supported by literature.

2 Theory

This chapter is the result of a literature review and will explain some of the core concepts used in the rest of the thesis: Digital platforms, Data marketplaces, Willingness to share data, Control, Trust and Risk. Explaining what has been researched on it so far, where the gaps in the literature are and how these concepts interrelate. Once we have a clear picture of all the concepts I will define a conceptual model. This model consists of variables and relations between the variables. I will define these relations as hypotheses that will be tested in later chapters.

2.1 Data marketplaces as a digital platform

Data marketplaces can be seen as a type of digital platform. A technical definition of a digital platform is given by Tiwana et al. (2010, p.676): "The extensible codebase of a software-based system that provides core functionality shared by the modules that interoperate with it and the interfaces through which they interoperate". This is similar to non digital platforms, that can be defined as a system with stable core components and variable peripheral components (Baldwin & Woodard, 2009, p.24). Digital platforms have also received attention from economic scholars that study them as multi-sided markets that leverage economies of scale by using network effects (Rochet & Tirole, 2003). This means that the more users a platform has, the more value it creates for the users. Creating a positive feedback loop that can result in a winner takes all scenario (T. Eisenmann et al., 2006). Network effects can be split into direct and indirect network effects. With indirect network effects, the value of a platform of one users increases with the amount of other users, for example the value of a data platform increases for a data consumer if there are more data suppliers. Direct network effects increase value for the same group of users, for example the value of a telephone increases for each user the more people use a telephone. Both the technical perspective with a stable core and variable peripheral components and the economic perspective are important to understand the workings of a digital platform. Aiming to bridge the gap between the technical and economic perspectives Gawer (2014) differentiates between platforms used within firms, within supply chains and across an industry, with each having a varying level of openness. A key factor for business to business platforms like data marketplaces is openness. Openness relates to the "Easing of restrictions on the use, development and commercialization of a technology" (Boudreau, 2010, p.1851). Generally there are two ways of regulating openness of a digital platform: By providing open access (and reduce control) or give partial access through boundary resources like API's (Karhu et al., 2018). An open platform should in theory be able to attract more users, but allow for less control by the platform owner (T. R. Eisenmann et al., 2009). This openness will be a key concept in data marketplaces that will be described later.

Data marketplaces are multi-sided platforms that match data providers and buyers (Abbas et al., 2021, p.3321). They share a lot of properties with digital platforms, but because they have data as their main product, they should be treated differently. Data can be de-contextualized, combined with other data and re-used in other contexts (Aaltonen et al., 2021). Data also needs this context to be valuable. Because of this data marketplaces typically need to be more sector specific than other digital platforms. Leading to a more fragmented market and less network effects (Mosterd et al., 2021). This is different from the standard winner-takes all scenarios of typical digital platforms. Data sharing also brings complicated legal implications to companies. The General Data Protection Regulation (GDPR) says that companies need to be clear about what they collect data for and that a person should be able to have their data removed (European Commission, 2016). This can become a problem if a company decides to sell that data. While the exact legal implications of this are out of the scope of this research. The implications it brings to the design and use of data marketplaces are relevant.

Important topics from the literature of digital platforms such as openness (De Reuver et al., 2018; Broekhuizen et al., 2021), governance and control mechanisms (Tiwana, 2014) are very relevant for the adoption of data marketplaces. A more open data marketplace should (like a more open platform) attract more users. Understanding digital platforms, helps understand data marketplaces better. However, it's important to note that they should be considered a special type of digital platform, because they have data as their main product. This brings many complications as just explained.

2.2 Organization's willingness to share data

The commercialization of data marketplaces is lacking (Spiekermann, 2019; Abbas et al., 2021). This thesis aims to create a better understanding on why this is the case. Before looking specifically at data marketplaces it's relevant to look at literature about the adoption of other B2B data sharing systems. Several factors are identified to influence this adoption:

Perceived benefits is perhaps the most important factor when looking at adoption of data sharing systems. Lee et al. (2000) investigated this for sharing data within supply chains and found that the potential value is very high. Li et al. (2006) also found that there were many benefits to sharing data between businesses if coordinated correctly. Chwelos et al. (2001) used a survey and found that along with the perceived benefits, readiness and external pressure were also good predictors of willingness to adopt a data sharing system.

Organizational readiness in the context of B2B data sharing can be split into many sub factors, with the most important ones being: IT staff skills, economic health and application interoperability (Mouzakitis & Askounis, 2010). Kuan and Chau (2001) used the Technology-Organization-Environment framework (Tornatzky & Fleischer, 1990) to classify different factors for the adoption of data sharing systems in SMEs. They also found that perceived direct benefits, perceived costs (readiness), perceived technical competence (readiness) and perceived industry and government pressure were important predictors of technology adoption.

Trust and relationships between companies are another way to look at willingness to share data. Chen et al. (2014) looked at knowledge sharing through this lens and found that factors such as shared goals, social relational embeddedness and influence strategies lead to more trust, collaboration and knowledge sharing. Ramadhan and Samadhi (2016, p.859) called trust ”a catalyst that facilitates strategic business interactions and knowledge sharing among independent firms”. Aiming to identify prerequisites for sharing industrial data Müller et al. (2020) performed a large study, combining expert opinions and empirical data. They also found that trust, social interaction and benefit sharing were some of the most important factors.

In addition to the research in these three important factors, there have been studies aiming to identify what makes a company willing to share data in various contexts (Penttinen et al., 2018; Kazantsev et al., 2022; Fu et al., 2014). They also identified ease of use, reliability and security as important factors. When looking at case studies perceived value, trust and readiness were also some of the most critical success factors (Holler et al., 2019; Dwaikat et al., 2018).

2.3 Sharing data in a marketplace

Data marketplaces work differently from data sharing platforms studied in the previous paragraph. Jovanovic et al. (2021) defined three types of industrial B2B platforms. According to their framework data marketplaces can be considered platform ecosystems. Which is different than the literature in the previous paragraph, covering mostly supply chain platforms. There are many differences between these types of platforms in their architecture, governance and services. So far there has not been a lot of research into how this affects the adoption. In fact, there have only been a handful of studies into the services, adoption and use cases of data marketplaces (Abbas et al., 2021). De Prieelle et al. (2022) did research on the adoption of data marketplaces in the horticulture industry. They found a lot of the same factors as described in 2.2 such as: perceived benefits, technological and human readiness, ease of use and trust in other users. It would seem logical that the results are similar, however I expect that companies would look differently at risk, trust and control when sharing in a marketplace.

Where the supply chain and ecosystem platforms differ mostly, is in their openness (Jovanovic et al., 2021). Data marketplaces are more open, which allows for more parties to share their data, at the cost of control (Richter and Slowinski, 2019;Koutroumpis et al., 2020). This creates a different setting for sharing data than with business partner. Through marketplaces companies would share data with more companies that they would trust less (Richter & Slowinski, 2019).

2.3.1 Risk

Data has some unique properties that makes trading it very different then other goods. As shown earlier, raw data can be used and reused in many different contexts (Aaltonen et al., 2021). A known example of this is accessing a website. When accessing, the website needs certain data from your computer such as your IP address. This data can be stored using cookies and reused for advertising purposes. While the original reason you send this data to the website is to request a web page. This is what makes sharing data valuable, but also unpredictable. A company can sell its data so it can generate value in many other areas, but they can't know exactly in what ways the other party will use it. I identify two major risks associated with this: Sharing personal information and sharing intellectual property that could give a competitive advantage to a rival (Reimsbach-Kounatze, 2021). The first risk could bring convoluted legal implications of which companies are very reluctant. It becomes unclear who is the owner of the data once it has been shared. Even when the data is anonymized when sharing, data analysis could still reveal some sensitive data (Narayanan & Shmatikov, 2008). The second risk will perhaps make companies even more reluctant to share data. Data could reveal information about how the company handles their business process or could even reveal some company secrets or unpatented designs. The consequences of data misuse depend heavily on the context and type of data. For example medical data will have a very high risk associated with it due to being sensitive personal data (Azarm-Daigle et al., 2015). Sharing data in the automotive industry will carry other risks due to it being a very competitive environment and every new design feature that gets leaked can give an advantage to a competitor (Kerber, 2018).

2.3.2 Trust

Trust is a concept linked directly to risk. This apparent in the definition by Mayer et al. (1995, p.712): "Willingness of a party to be vulnerable to the actions of another party...". When sharing data in a marketplace there are two parties that a data owner should trust: The data consumer and the platform. Inter organizational trust along with other relational factors have been studied extensively in relation to sharing data (Chen et al., 2014; Ramadhan and Samadhi, 2016). It is one of the key factors required for a company to share data (Müller et al., 2020). Most of the studies on willingness to share data have however considered trust in a setting between business partners, mostly within supply chains. Data marketplaces change this by allowing data owners to share data to many more parties. The lack of trust between owner and consumer can become a barrier for sharing this information (Christidis et al., 2022). There are mechanisms of increasing trust in data marketplaces such as reputation systems that determines access to the data (Roman & Stefano, 2016), and even systems that monitor the use of the data (Noorian et al., 2014). A lack of trust can also be compensated by increasing the control a data owner has, as will be described in chapter 2.3.3. However the trust in the control mechanisms and the platform owner does need to increase for this to be viable. This is shown in the study by Agahari et al. (2022) where respondents required more trust in the technology where there was less trust in other users. This is often the case when trustless technologies such as blockchain are used (Christidis et al., 2022).

2.3.3 Control

With a lack of trust, control can help reduce risk. In platform governance literature control refers to: "formal and informal mechanisms implemented by a platform owner to encourage desirable behaviors by module developers, and vice versa" (Tiwana et al., 2010, p. 680). In this study I will look at control from the perspective of data providers. Similar to Agahari et al. (2022) and Otto et al. (2019) I define control as the mechanisms data providers can use to determine data usage. This includes mechanisms that help providers determine who the data is sold to, what data is shared, for what purpose the data may be used and how the data flow is handled. What mechanisms are available to the data provider is determined by the platform owner.

A platform can have objectively more control through these mechanisms. However what really is a driving factor when it comes to the willingness to share data is the perceived control. If a person perceives they have control over the data, they will likely experience less risk in sharing this data. Increasing the actual control should help decrease the risk of misuse, but only if the users also perceive more control. This way having more control will help reduce risk of misuse, but will make the platform less open. A more open platform should attract more users (T. R. Eisenmann et al., 2009). This creates a balance between openness and control mechanisms that will be different in every setting based on factors like the type of data being shared and the trust among other users (Reimsbach-Kounatze, 2021).

Control mechanisms are typically split into formal and informal mechanisms. Formal mechanisms include: input control, process control and output control (Tiwana, 2014). Input control or gatekeeping include all methods of limiting access. This can be access to the platform itself or specific data sold on the platform. Process control are mechanisms that platform owners use to guide users into the desired direction. For example a platform can require users to submit data in an a certain format or have the data automatically be encrypted when uploading. Output control mechanisms work based on metrics of the data shared on the platform. For example data consumers could rate the data being shared on usefulness and privacy sensitivity. If a dataset is deemed to have sensitive data or isn't very useful it can be removed. Relation or clan control is the only type of informal mechanism identified in the works by Tiwana (2014). It works on developing shared norms and goals that make sure every user is contributing in a positive way to the platform. Relational control can be seen as a platforms mechanisms to increase trust between users. These different categories of mechanisms don't work in isolation, they often complement each-other.

Next to the classification of control mechanisms just discussed, I classify control mechanisms in another way: institutional and technical. Most data exchange between business rely on some sort of a mutual agreement. I consider this a set of institutional mechanisms. This agreement can determine what data will be shared, for what purpose it may be used and what purposes it may not be used. This can be in the form of a contract between the two parties, accepting a user agreement or simply the way that a marketplace or data exchange software handles its transactions. Traditionally this set of rules is created by both the data provider and consumer, who will share data only after agreeing. In a data marketplace this set of rules is determined by the data provider, but is dependent on the options provided by the platform. A platform might for example provide different types of access control options: provide open access, provide access to prescreened users, provide access to members of a specific group or provide access only to consumers approved by the provider. The platform might then also give options for controlling how the consumer can use the data.

In a perfect world all the rules in the usage agreement or contract should give the data provider all the control they need, however there are several flaws in this system. The data consumer might not read or understand the complete usage agreement and unintentionally misuse the data. The owner might not be able to cover all potential misuses in the agreement. The consumer could also intentionally misuse the data. In all these cases there is nothing physically stopping misuse from happening (Reimsbach-Kounatze, 2021) The only thing the owner could do to stop misuse after sharing data is to take legal action, which obviously is not preferable. Because of this there are technological mechanisms that help that the data is being used as agreed and give more control to the data owner. Some technological control mechanisms mentioned in the literature are:

- **Blockchain** has received a lot of attention in the data marketplace literature (Bajoudah et al., 2019;Christidis et al., 2022;Abbas et al., 2021). Its decentralized nature makes data sharing less reliant on trust in the platform owner and provides a secure way of creating transactions (Weber et al., 2016). Encryption of data and data transactions allow for better security. And lastly smart contracts allow the provider to set many rules to how the data is used and shared. These rules will be automatically executed. This is a perfect example of technology enforcing institutional control mechanisms.

- **Multi Party Computation (MPC)** is a combination of a decentralized marketplace, encryption and data analytics. It gathers data from multiple providers, combines them, runs data analytics and sends results to the consumer. The advantage of this is that the data providers can keep sovereignty of their sensitive data, because only the output of the data analytics is shared. This makes it excellent for sharing personal information while keeping anonymity(More & Alber, 2022)¿ But also makes it very good for sharing business data without disclosing anything sensitive, as is shown in the study by Agahari et al. (2022).

- **Standardization** in the way that data is shared is a form of process control. Figueredo et al. (2022) show that using an open standard (oneM2M) can help increase control by only allowing IoT data to be shared in a specific manner. A different way of standardization is through principles such as the FAIR principles. Calancea and Alboaie (2021) used these principles to control B2B data sharing, also making use of their own decentralized, blockchain based architecture.

- **Reputation systems** are technical mechanisms that help build trust between the users (Tadelis, 2016). While they are prone to have some bias, they can still be useful for identifying the quality of a user or dataset. There have been some examples of reputation systems in data marketplaces (Roman & Stefano, 2016). Even some where the reputation system is linked to monitoring software that tracks if the data is used as intended (Noorian et al., 2014).

Willingness to share data is a complicated matter. It can be looked at from many perspectives, which all have a contribution a companies decision. This research will contribute to this by focusing on the perceived risks of sharing data. This risk can be mitigated by creating trust between companies or increasing the companies perceived control. In data marketplaces there is likely not that much trust, which could mean that control becomes a more important factor. In the next chapter I will go into more detail into the relationships between these factors.

2.4 Relations between control, trust and risk

Now that the concepts are defined and given a theoretical background it's time to look closer at the relationships between the concepts. I will do this by creating hypothesis that describe the relationships. These hypotheses will later be tested as described in chapter 3. The combination of these hypotheses create a conceptual model. This model is visualized through a causal diagram in figure 2.

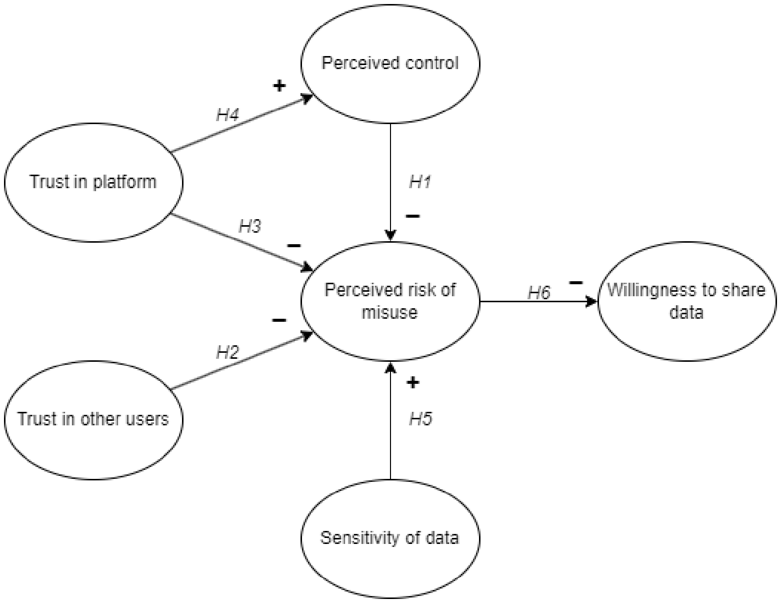


Figure 2: Causal diagram

In the previous section I explained how control mechanisms should limit the ways in which data can be misused. Rules set by data owner and the platform limit the ways in which the consumer is allowed to use the data and technical control mechanisms help enforce these rules. If the data provider perceives that these mechanisms provide enough control, they would also perceive lower risk of data misuse (Reimsbach-Kounatze, 2021; Agahari et al., 2022; Richter and Slowinski, 2019) Therefor the first hypothesis is:

H1: Higher perceived control will lower the perceived risk of data misuse

As described in 2.3.2 trust is defined as the willingness to be vulnerable to the actions of another party. In other words if a company trusts the other users of a marketplace they perceive less risk that their data will be misused (Agahari et al., 2022; Kazantsev et al., 2022; Richter and Slowinski, 2019; Roman and Stefano, 2016; Beldad et al., 2010). In data marketplaces however I assume that the trust is lower than in traditional data sharing. I therefor expect the following relationship:

H2: Higher trust between users will decrease the perceived risk of data misuse

Similar to the trust in other users, the data owner should trust the provider of the platform. If the platform is trustworthy, the data owner would assume that the risk of data misuse is lower (Agahari et al., 2022;De Prieelle et al., 2022; Beldad et al., 2010). Decentralized platforms have no clear owner. In this case I consider the trust in the platform to be the trust in the technology and the decentralized structure as a whole to be the trust in the platform, instead of trust in the platform owner (González et al., 2023; Agahari et al., 2022). In either cases I expect that:

H3: Higher trust in the platform will decrease the perceived risk of data misuse

The more companies rely on 3rd parties (platform owner) or technology (platform) for control over their data, the higher their trust in these 3rd parties and technology needs to be. If there was no trust in the platform, it doesn't really matter how much control they provide. Therefor I argue that higher trust in the platform will result in higher perceived control (Agahari et al., 2022; De Prieelle et al., 2022), as described in the following hypothesis:

H4: Higher trust in the platform will correlate with higher perceived control

A variable I'm less interested in, but I expect will play a large role in determining perceived risk, is the sensitivity of the data. This is determined by many contextual factors such as what sector they work and how much of a negative impact potential data misuse can have(Reimsbach-Kounatze, 2021; Martin et al., 2017; González et al., 2023). I therefor state that.

H5: The more sensitive the data is, the higher the perceived risk of misuse

There are many factors impacting how willing companies are to share data, as is clear from 2.2 and 2.3. Out of the variables that I'm interested in I argue that only perceived risk of misuse will directly influence a companies willingness to share data. Both perceived control and trust are important for the willingness to share data, but only because they can reduce the risk. The goal of this research is to find out better why companies are reluctant to share data. To test if risk has a meaningful contribution to this, I will test if:

H6: Higher risk of misuse will correlate with less willingness to share data

Other factors influencing the willingness to share data such as the perceived value and ease of use are not part of this study. This conceptual model gives an overview of the relationships between the factors I'm interested in studying. The next chapter will describe how I aim to validate this model.

3 Methodology

Now that I've identified the constructs in the literature and created a conceptual model, I want to test this. This chapter will describe how the data was gathered for validating the conceptual model. I'll do this by describing the data collection process, how the questionnaire was designed and how the latent constructs are operationalized.

To validate the model described in chapter 2.4 I want to use covariance based structural equation modelling. Covariance based structural equation modelling is perfectly suited for testing the hypotheses that describe relationships between latent constructs (Hair et al., 2021).

I wanted to select a target group that would consider sharing their data to be risky. If the respondents wouldn't think the data was very sensitive and don't see the risks in sharing it, it is unlikely we see how control and trust impacts this. My focus is on sharing business data, what makes this different from sharing individuals data is that business data can be sensitive for other reasons, such as giving away a competitive advantage. However, business data can still contain personal data. For this research it would be undesirable if there was a lot of personal data in the business data, as there would be many other factors involved in a respondents perceived risk and willingness to share data. For example if I asked respondents about sharing business data containing lots of personal information they would consider the legality of sharing the data. I opted to select internet of things (IoT) data as the data to be shared and the manufacturing industry as the target group. So far industry only has a very small marketshare, when comparing it to other paid data marketplaces (Andres & Laoutaris, 2022). As mentioned in the introduction IoT data from manufacturing companies is still very underutilized (Frank et al., 2019) and sharing this data could help improve this (Dai et al., 2020). Industrial IoT data as it's called, consists of data from sensors in production processes as well as other data about the production process (e.g. test reports). A collection of this data can contain both the live status of a factory and historical data about the production. This data can be very helpful for other companies using similar machines or production processes. However, companies will be reluctant to share the data as this can give away a lot of information that can help a competitor. It's unlikely that it has a lot of personal data, making it suitable as a target scenario for the survey.

The population I want to describe in this research is companies in the manufacturing industry that can adopt a data marketplace. A representative sample would include respondents that have influence over the adoption of data marketplaces in a manufacturing company. This can be someone in various roles, but needs to have some decision power. Ideally they would already be very knowledgeable on IoT and data marketplaces. The manufacturing industry is a large and diverse population. For my research to be representative I would need to have respondents to be spread across many different sectors within the industry.

3.1 Ethic approval

The Human Research and Ethics Committee of TU Delft has officially approved the survey on the 10th of February 2023.

3.2 Data collection

Data was collected using the online survey platform Qualtrics. I were aiming to get 300 responses. Respondents for the survey were recruited using the Prolific platform. An online platform that allows researchers to find respondents and respondents to receive a small financial reward for completing surveys. Prolific asks its users to give some personal information that researchers can use as filters. I've opted to use the filters: "Which of the following best describes your role at work?" and "Which of the following categories best describes the industry you primarily work in (regardless of your actual position)?". From the industry role filter I allowed: "Upper Management, "Middle Management", "Junior management", "Consultant" and "Researcher". From the industry filter I allowed: "Manufacturing" and "Other Manufacturing". I also selected that respondents need to reside in the EU, UK or USA. I payed each respondent £2,60 for the 15 minute survey. This is an hourly rate of £10,40 which is slightly higher than the recommended £9,00/hour. This hourly rate was chosen because I only had 829 potential respondents, out of 120.000+ potential respondents using the platform. The rewards were paid for by a leftover budget from earlier studies by Wirawan Agahari, who looked into Multi-Party Computation, a technology related to data marketplaces. Paying respondents will create some bias in the results and incentives for the respondents to just finish the survey as quickly as possible. How I deal with this will be described in the results and discussion.

The collection of data happened very quickly. The survey was published on the afternoon of Friday 10 March and I had 299 respondents within one day. The last few responses were locked out of the survey because Prolific paused the response collection at 299 respondents. I restarted collecting data to get the last respondent on Sunday 12 March. This gave us 300 complete responses and 30 uncompleted responses.

3.3 Sample description

Out of the 330 responses I've received, 30 did not complete the survey after the comprehension check. I expected some respondents to fill out the survey too quickly, because they only wanted the reward money. To filter these responses I looked at the completion time. After plotting the distribution of completion time I decided 3 minutes was a good cutoff value for a valid response. This gave us a total of 295 valid responses, with two responses missing one item. Distributions that describe the final sample can be seen in figure 3. The complete aggregated demographic data can be found in appendix D.

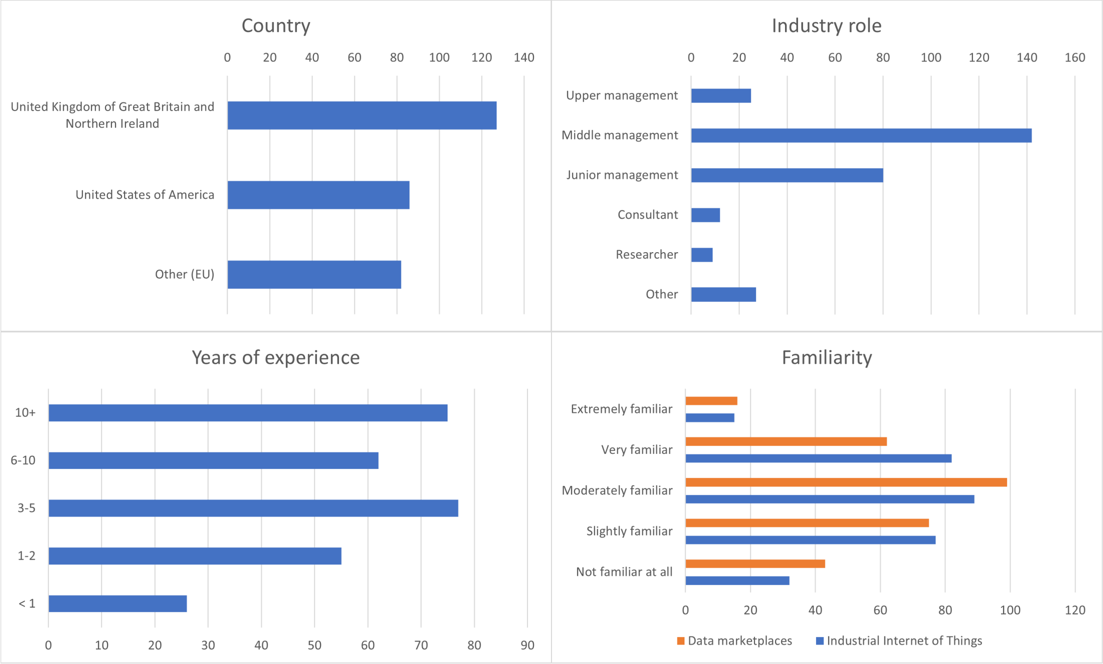


Figure 3: Demographic description

The final sample of 295 had a lot of respondents from the USA and UK (combined 72%). This has to do with the number of Prolific users being much higher in these countries. This makes the results representative for western countries, but might give a little bias because the cultural differences of USA and UK compared to EU countries. Most respondents have a role in middle or junior management (combined 75%). I expect that these people have a lot of influence in the adoption of data marketplaces. This makes them very representative when it comes to describing a companies willingness to share data. The upper management group however would likely have an even bigger influence and only has 8% representation. There's also a small group (9%) that entered "Other" as an industry role. These people would have previously selected one of the industry roles within my criteria in Prolific. I therefor assume that these people were previously in a position where they had influence over a companies willingness to share data. However it remains a bit uncertain how representative these responses are. The responses had people with many different years of experience. A small number of respondents had less than a year experience in their industry role, with the rest being quite evenly distributed between 1-2, 3-5, 6-10 and 10+ years. This is very representative for the overall population of people that have influence over the adoption of data marketplaces. Data marketplaces are a relatively new concept. It's likely that people with knowledge of data marketplaces would have the most influence about the adoption. However data marketplaces are such a new concept, especially in the manufacturing industry. A lot of people that would be in a position to adopt a data marketplace in the future might not have heard about it yet. The responses actually showed that there were quite a few people already familiar with data marketplaces (34% moderately familiar and 26% very and extremely familiar). For a survey done today I think this will be very representative for gaining insight on the willingness to share data. It would be interesting if the results would change if the survey was done again in the future, when data marketplaces are more common. Familiarity with industrial IoT was a bit higher than familiarity with data marketplaces, however it was lower than I personally expected. I think it would be difficult to judge the risks of sharing data when you don't understand what data it is you're sharing. This does make the sample a bit less representative for the overall population. Still, only 11% said they were not familiar at all with Industrial IoT, which is a small portion of the full sample.

Ideally the sample would have respondents that supply to an even spread of many different industries. Because I could not know which sectors the respondents supply to and because there are so many possible sectors, I included an open text question. After collecting the responses, I scanned the answers and categorized them, the results can be found in appendix D. There were 5 sectors that were common in the answers: Manufacturing, Food, Medical, Consumers/Retail, Automotive. Manufacturing was by far the most common (40%). This could be because some people might have interpreted the question as" which sector do you work in?". It could also just be that a lot of the respondents are a manufacturing company, supplying to other manufacturing companies. Besides that there was a good spread across many different industries, with many responses falling under the other category, which included many other sectors such as: Chemical, Rail, Aviation and many more.

Overall I think the sample is very representative for people that are now in a position to adopt a data marketplace. They should be able to form an opinion on the perceived control, trust and risk when sharing data in a marketplace. The sample also has a good spread across different industries, level of experience and industry roles. The familiarity with IoT and data marketplaces is not that high. This is not unexpected as they are still relatively new concepts, but will make the sample a bit less representative. As in reality someone would do research on these topics before deciding to share data or not.

3.4 Questionnaire design

The goal of the questionnaire is to measure the relation between the 6 constructs from the causal model in figure 2: Trust in platform, Trust in other users, Perceived control, Sensitivity of data, Perceived risk of misuse and Willingness to share data. In order to do that reliably I need to first give the respondents some information on data marketplaces. This chapter will explain how the survey is put together, followed by a chapter explaining how the constructs are measured. Figure 4 gives an overview of the design.

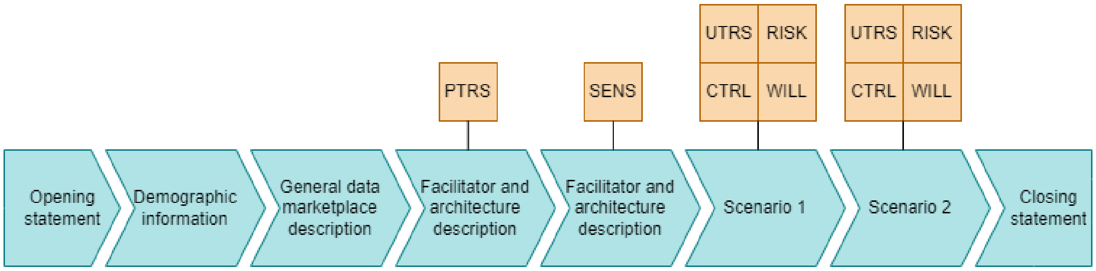


Figure 4: Questionnaire design

3.4.1 Demographic information

After the opening statement is read and agreed to, the respondents are asked for some personal information. This includes a multiple choice question for industry role, years of experience, country of residence and an open field for sector they manufacture products for. These serve to describe the demographic and it will not be possible to identify a person based on this information. The respondent will then be asked how familiar they are with Industrial internet of things and Sharing data on a marketplace. This is also to describe the demographic.

3.4.2 Marketplace description

Before asking the respondents about the constructs they receive an explanation of a data marketplace, found in appendix B.2 This is so that the respondents are on the same page about what a data marketplace is, but also to exclude constructs that are not included in this research. Starting the description the respondents are given a scenario that makes it very easy for them to share data. In reality many manufacturing companies do not yet collect that much data (Frank et al., 2019). They are told that their company has spent the last 5-10 years investing in smart manufacturing systems and they have access to all data that can be generated by their machines and manufacturing processes. One of the key factors for adopting data sharing systems is organizational readiness (Kuan & Chau, 2001). To make sure this is not influencing the response I state that all their systems are inter operable, everything can get accessed easily and it takes little to no effort to train personnel to use the marketplace. I tell the respondents that sharing this data is almost as easy as clicking a ”Share data” button. Pricing is another hurdle that has received a lot of attention from researchers (Abbas et al., 2021;Mao et al., 2019). I tell the respondents they will receive a ”fair price”, which is enough to cover any expenses that they need to make to share the data and some profit, but not enough profit to change their business model to share more data.

3.4.3 Marketplace facilitator and architecture

How the data is stored and who is facilitating the data exchange is very important. If data providers (or consumers) don’t trust the platform, they won’t use it (Chang et al., 2020). There are many different types of parties that could create a data marketplace and even more configurations of how they can structure this marketplace. I considered 3 types of facilitators: A large known company that launches a marketplace selling their data and allowing others to sell on it as well. A small independent company that purely focuses on facilitating data exchange and a completely decentralized marketplace. I opted for the small independent company, because I think it would be the most trustworthy for a manufacturing company. One from a large company would rely on the respondents trust of that company or large companies in general. A completely decentralized marketplace would rely more on trust in technology and trust in many unknown parties. A small independent company is something that a person working in the manufacturing industry can easily understand and would expect to have the least motive to misuse their data.

As described in the literature review by Abbas et al. (2021), there are a lot of papers about the architecture of data marketplaces. Nearly all of them are using some sort of decentralized mechanisms either blockchain for storing transactions (González et al., 2023), decentralized databases for storing data or privacy enhancing mechanisms such as MPC for sharing computational results. Because of all this attention and of course its properties for trustless transactions (González et al., 2023), I wanted to include blockchain as a way of storing the transactions done in the marketplace. There are a few papers examining architecture options for IoT data marketplaces, though none specifically for the manufacturing industry. Sharing IoT data is slightly different than sharing other data, mainly because it includes live data, instead of only static data (Misura & Zagar, 2016). Out of the available descriptions I chose to adapt the marketplace description by Gupta et al. (2021), but simplify it so the respondents can understand it better.

This architecture consists of 3 tiers: The participants, the facilitator and the regulators. The participants are the data providers and consumers, they can sell data to each other. Every transaction that they make, goes through the facilitator. This facilitator is the small independent company described earlier. They will have geographically distributed locations facilitate data exchange within and to specific service areas. How the data is handled depends on the marketplace configuration I will discuss later. In some cases the data is stored on a decentralized marketplace, managed by the facilitator, in other cases the consumer has direct access to an API of the provider and in other cases the data is send through the facilitator. In all these cases the facilitator will log every transaction on a blockchain. In the case of direct access to the API the facilitator will log the fact that the consumer has access to the API and the agreed terms of use. The regulators are government bodies of various countries that make sure everything is compliant with local privacy laws. They have access to the transaction blockchain and will make sure no privacy sensitive data is logged on the blockchain.

Along with a description of the facilitator and the platform architecture I show the respondents a simplified version of the 3-tier system architecture of Gupta et al. (2021). The simplified version is shown in figure 5. Appendix B.3 shows the description included in the survey.

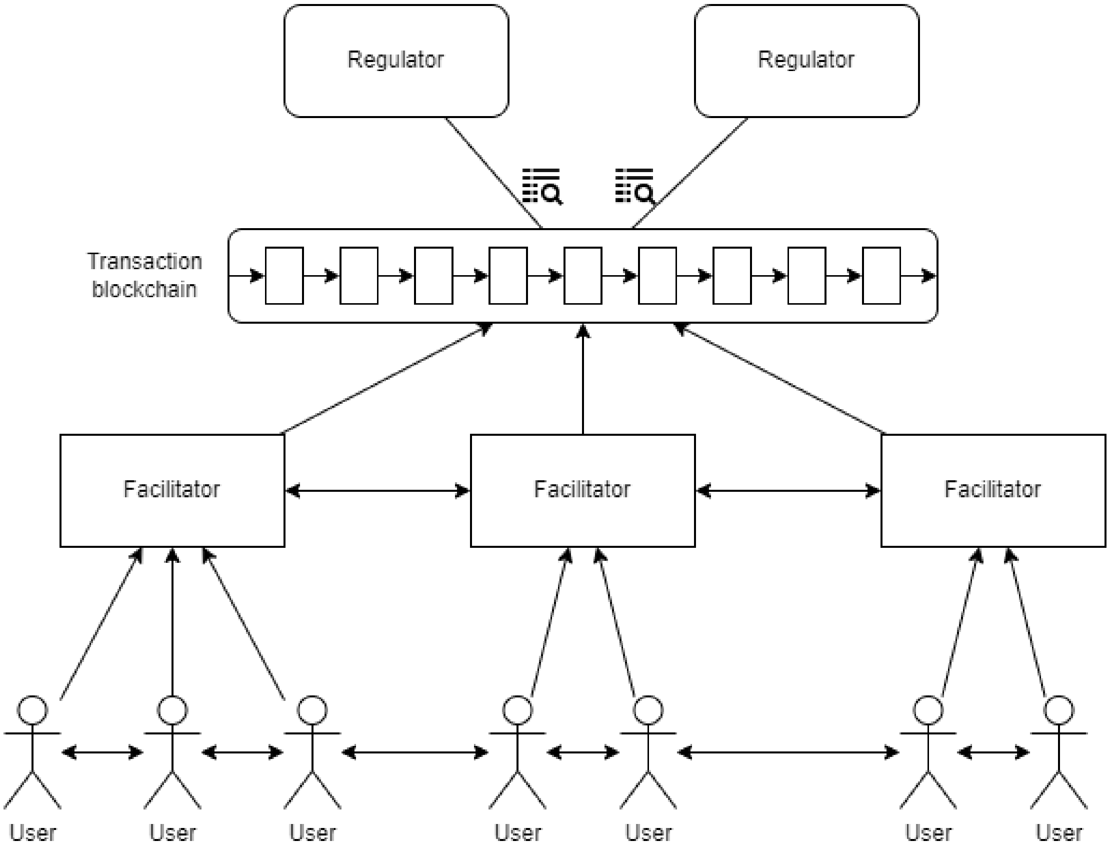


Figure 5: Simplified version of 3-tier system architecture, adapted from Gupta et al. (2021)

3.4.4 Scenarios with different control mechanisms

As mentioned about the architecture there is no standard design or golden formula for a data marketplace. Especially for IoT and Industrial IoT there is still a lot of uncertainty about how the marketplaces will function. The theory chapter also showed many different control mechanisms. To get a better sense of how much perceived control and perceived trust these mechanisms give, and because there is no single design that I can comfortably say is representative for industrial data marketplaces, I want to test multiple control mechanisms. I do this in the survey by giving the respondents multiple scenarios with different sets of control mechanisms. The control mechanisms are split into two parts: Signing up and Data exchange. Firstly there were two options for trust and four for control, creating 8 scenarios. The survey would randomly give the participants 4 out of 8 scenarios. Testing this proved the survey was too long and participants were no longer reading the scenario descriptions. It was therefor shortened to only two scenarios: a low trust, low control one and a high trust, high control one. Why I selected these scenarios over others will be explained at the end of this paragraph.

As mentioned the signing up description has two options. I expect that they give different levels of trust to the data consumer. The first option is a very open marketplace, which allows everyone to sign up with just an email address. I wanted to include this, because it measures how people feel about a truly open marketplace. Where people won't trust each other, but have to rely on the control mechanisms to minimize risk. The second option includes a verification process as described by Cárdenas and Molano (2022). In this option every user needs to go through a verification process when they sign up. This way the other companies can at least be sure they know who they're selling the data to, which should give more trust.

For control I also ended up with two options. Starting with an open marketplace where the data is simply listed, stored in a decentralized database and can just be downloaded by anyone agreeing to the terms of use and paying the price. The second option should give the respondents the more control. It allows the data provider and consumer to negotiate over the terms of use. Once everything is agreed to they will create a smart contract that will automatically handle all the transactions.

After each scenario I will ask the respondents about their perceived control, trust in other users, risk and willingness to share data. The trust in the platform and data sensitivity will only be measured once, because the facilitator, architecture and the data they can share will stay the same across the scenarios. I did consider including different options for the platform architecture, but the option by Gupta et al. (2021) seems convincing and including multiple architecture descriptions would have made the survey more complicated for the respondent. appendix B shows the description of both scenarios.

There are a few reasons why I chose these two scenarios: Firstly the low trust, low control scenario offered the respondents a truly open marketplace. This is in line with the digital platforms literature that suggests that a more open marketplace will attract the most users (T. R. Eisenmann et al., 2009). Measuring people's perceived control, trust and risk in this scenario gives a sort of baseline response and shows how willing companies would be to share data when the risks are higher. Secondly the high trust, high control scenario offered the respondents much more control. Seeing how they perceive this helps validate the effectiveness of validating users and smart contract use in increasing the perceived control and risk. It also allows us to see if increasing the control and trust mechanisms really does help reduce the perceived risk (by increasing the perceived control and trust). Comparing the two scenarios I expect that there will be a lot more willingness to share data in the second. If this is the case the results would show how willing companies are in a open marketplace and how willing they are if they perceived more control.

3.4.5 Trial run

As implied by the previous chapter I performed a trial run of the survey before sending it. The trial version was completed by 6 people: 3 Fellow researchers familiar with data marketplaces, 2 people with a social sciences background to help with general questionnaire design and 1 director of a medium sized manufacturing company. While I did approach more people in the manufacturing industry, I was unable to get a response from them in time. The trial run helped fine tuning the survey, getting the wording correct and getting details right such as including a progress bar and adding a comprehension check. Every respondent in the trial version mentioned that the survey had too much text and took longer than the stated 10 minutes. To improve this I trimmed unnecessary text, only included 2 scenarios instead of 4, added bold text to emphasize differences in the scenarios and increased the stated time from 10 to 15 minutes.

3.5 Measuring perceived control, trust and risk

In this section I will explain how each construct is operationalized. The exact wording of the items, along with the variable names can be found in appendix C.

3.5.1 Platform trust - PTRS

Trust is a complex construct with many different possible approaches to measure. As is clear by the extensive literature review on antecedents of online trust by Beldad et al. (2010). Which already reveals too many variables to include in a simple survey. I am interested in measuring the trustworthiness of two parties: The platform and the data consumer. Some people are more trusting than others, which will always create a variance in the data. Some people or organizations are more trustworthy than others and how the platform and data consumer can present themselves can influence their trustworthiness (Beldad et al., 2010). According to Mayer et al. (1995) Ability, Benevolence and Integrity are the main factors of perceived trustworthiness. So when measuring the trust in the platform and trust in the data consumer I will measure if the party seems to: Be capable of protecting the data and not allowing misuse, has the intention to not abuse the data and is honest about their intentions.

Trust in the platform for us means two things: Trust that the facilitator is not likely to misuse the data. This comes down to the integrity and benevolence of the facilitator. The other part is the trust in that the facilitator has the ability to keep the data from being misused. Which has to do with the architecture of the platform and the respondents trust in the technology and the blockchain network (Weber et al., 2016). To operationalize this I adapted 4 questionnaire items from Chang et al. (2020). They were used in a questionnaire about trust in online B2B marketplaces for items such as LCD monitors. The items measure the trust in the intermediary such as the website that facilitates this exchange. While the mechanisms a regular marketplace uses to create trust may be different than a data marketplace, I think the measurement of trust in this intermediary is very similar to the trust in a data marketplace platform.

3.5.2 Sensitivity of data - SENS

As mentioned earlier I expect the sensitivity of the data to have a large effect on the perceived risk. Someone who doesn't consider the data to be very sensitive is unlikely to see a large risk in sharing the data. Industrial IoT data has two components that can make them sensitive: Have personal information or give away a company advantage or even trade secrets (Reimsbach-Kounatze, 2021). Most measurement scales of data sensitivity only focus on the privacy aspect (Martin et al., 2017;Agahari and de Reuver, 2022), and they do this mostly from the perspective of a consumer giving away personal information to a business. I found the scale used by Kehr et al. (2015) the most useful. Which is a simple 7-Point scale from "Not sensitive at all" to "Very sensitive". Before asking the participant to rate the sensitivity I explain that they can exclude some critically sensitive information such as names of products or inventory numbers.

3.5.3 Trust in other users - UTRS

Similar to the trust in the platform, trust in other users can be measured in many different ways. I am interested in how likely the respondents think the other user will misuse their data. I measure this with the same three concepts as trust in platform: Integrity, Benevolence and

Ability (Mayer et al., 1995). To do this, I chose to use the same items as Agahari and de Reuver (2022). They use 3 items to measure the trust in buyers of a data marketplace using a Likert scale. The paper uses it slightly different, because it measures trust from the perspective of a consumer selling their data to businesses.

3.5.4 Perceived control - CTRL

Control over data is complicated, as is apparent from the chapter 2.3.3 where I list different ways of defining control. Control over data is different from control over other goods, because it's often unknown how the other party will use the data (Reimsbach-Kounatze, 2021). To measure the control data providers perceive over the data, I adopted items from Xu et al. (2008). They performed a large survey about the perception of privacy issues when sharing information online. Items such as: "I believe I have control over who can get access to my personal information collected by these websites" are very similar to what is needed to measure perceived control over business data.

3.5.5 Perceived risk - RISK

As with the perceived control and sensitivity of data most of the literature on perceived risk of data sharing is about releasing privacy sensitive data. I am interested in the risks of sharing privacy sensitive data, but also risk of sharing data that could give away a competitive advantage. While there has been plenty of research on trust and perceived risk when sharing information between business partners, such as Cheng et al. (2013), they focus on the relationship between companies that know each other. So instead I will adopt a measurement scale used for measuring the risk of sharing privacy sensitive data to measuring the risks sharing business data. With items such as: "The data could be inappropriately used". The scale is used by both Kehr et al. (2015) and Xu et al. (2008). I included the item "There would be a high potential for privacy loss..." twice, changing it to "There would be a high potential for giving away a competitive advantage" the second time.

3.5.6 Willingness to share data - WILL

After asking about the perceived risk, I will restate the assumptions mentioned at the start: Have enough data to share, data is ready to share, receiving a fair price and ability to remove sensitive data. This is so we can purely focus on the effect of risk on willingness to share data. Measuring the willingness to share data is done using the same Likert scale items as Agahari and de Reuver (2022).

All the to be measured constructs are operationalized and the survey is ready. In the next chapter the results of the survey are described.

4 Results

This chapter will give an overview of the results. It will do this by describing the demographic and what the data looks like. After that I will analyze the data using structural equation modelling. This chapter describe how well the model measures the constructs, how well the model fits the data and what the regression between the different constructs look like.

4.1 Data description

Before going into the measurement model and the relations between the constructs, I will first evaluate the absolute values of the data. The two scenarios described in chapter 3.4 measured 4 constructs twice, once in a low trust/low control scenario and once in a high trust/high control scenario. In the measurement model and structural model later I will look at these scenarios as two seperate models. These models use the same values for the items PTRST and SENS.

Table 1 shows the mean values of the constructs, by taking the mean of all the measured items belonging to each construct. Appendix E shows the mean, standard deviation, skewness and kurtosis statistics for all the measured items. In general we can see that the trust in the platform was considered quite high (5.10 on a scale of 1-7) and people considered the data to be sensitive as well (4.7 on a scale of 1-7). Looking at the differences between scenario 1 and 2 we can see that my assumptions were true. People reported much higher trust in other users and control in scenario 2, compared to 1. By just looking at the mean values you can also see that people also perceived less risk in scenario 2 and were more willing to share data. How well they correlate with each other will be described later in the structural model.

	Mean		
	Both models	**Model 1**	**Model 2**
PTRST	5,10		
SENS	4,70		
UTRST1		2,88	
CTRL1		3,09	
RISK1		4,93	
WILL1		3,83	
UTRST2			4,03
CTRL2			4,82
RISK2			4,06
WILL2			4,78

Table 1: Mean values, all values are on a 1-7 scale

Looking at the distributions of the data it was notable that some variables were not normally distributed. Figure 6 shows two examples of this. It looks like people are either trusting or not trusting and feel they have control or no control, there aren't many people choosing the middle option. We see similar distributions in the RISK1 and WILL1 items. This has some implications on the model fit which will be discussed in the chapter 4.3.

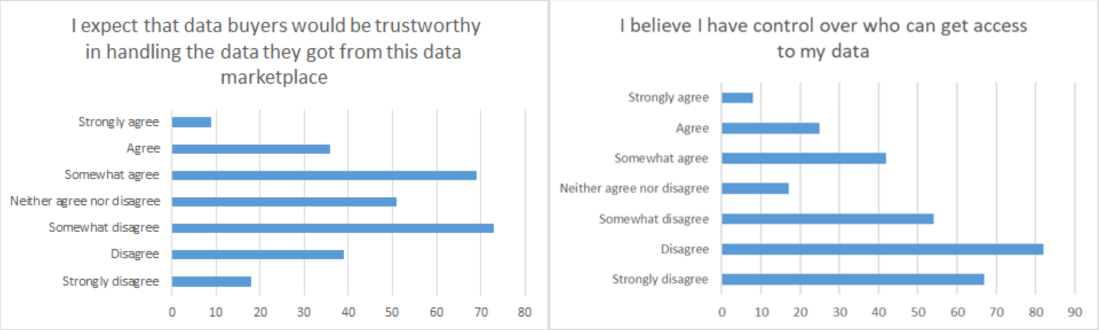


Figure 6: Distributions of UTRST1_1 and CTRL1_1 showing non normality

4.2 Measurement model

The measurement models help determine if the indicators are suitable for measuring the constructs. Later I will use these constructs to look into the relations between control, trust and risk. To do this I use confirmatory factor analysis. This is a type of Structural Equation Model that aims to define the variation and covariance between constructs and its indicators (Brown, 2015). I will determine if the indicators are a good fit by looking at the factor loadings, reliability and validity. All measurement models have been estimated using the maximum likelyhood algorithm. The software R and RStudio were used for estimation, using the Lavaan package. The R scripts of the measurement models can be found in appendix H

4.2.1 Loadings

Table 2 shows the factor loading of model 1. Hair et al. (2019) claims that indicators below a value of 0.708 do not adequately reflect the constructs, which reduces the reliability of the structural analysis later. I noticed the items CTRL1_2 and RISK1_2 have loadings of 0.507 and 0.626 respectively.
The low factor loadings of CTRL1_2 and RISK1_2 can also be explained by looking at the items. The item CTRL1_2 is about control over what data is released by the marketplace. In both scenarios I make it clear that the respondents themselves control what data they put on the marketplace. It's likely that respondents that did not experience a lot of control in the other items still felt like they had control over what data was shared, giving it a low correlation to the control construct. This is also reflected in the mean values of this item. The CTRL1_2 item had a mean of 4,044, while the other CTRL1 items had much lower means (2,973, 2,641, 2,722).

The item RISK1_2 is about the risk of giving away personal information. The other items are about general perceived risks about sharing data and risk of giving away a competitive advantage. The mean value of this item (4,295) is much lower than the other items (4,881, 5,075, 5,319, 5,064). This indicates that the respondents perceived lower risk of sharing personal information, compared to other risks.

I therefor chose to remove these two items. The model fit will later be assessed in more detail, but I could already notice a large improvement in the Chi square statistic, which dropped from 347 to 233 in the measurement model. The updated factor loadings of model 1 can be found in appendix F.2.

	PTRST	**UTRST1**	**CTRL1**	**RISK1**	**WILL1**	**SENS**
PTRST_1	0,722					
PTRST_2	0,86					
PTRST_3	0,897					
PTRST_4	0,776					
UTRST1_1		0,914				
UTRST1_2		0,952				
UTRST1_3		0,948				
CTRL1_1			0,812			
CTRL1_2			0,507			
CTRL1_3			0,915			
CTRL1_4			0,901			
RISK1_1				0,831		
RISK1_2				0,626		
RISK1_3				0,751		
RISK1_4				0,744		
RISK1_5				0,778		
WILL1_1					0,929	
WILL1_2					0,961	
WILL1_3					0,949	
SENS_1						1

Table 2: Model 1 standardized factor loadings

Model 2 will have the exact same items as model 1, the difference between the two is the scenario presented to the respondents beforehand. So I expect to see similar results in the measurement model. Looking at the loadings (shown in table 3) however, CTRL2_2 and RISK2_2 had much higher loadings of 0,775 and 0,783. The mean values of the items within these constructs does reveal some sort of explanation for this. The mean values of the other items within these constructs have gotten closer to the values of CTRL2_2 and RISK2_2. The overall control of the second scenario was higher and overall risk was lower. While the control over what data was released and risk of giving away personal information did not change as much. As an experiment I did remove the same items as measurement model 1. This resulted in a very large improvement of model fit. The Chi Square statistic dropped from 526 to 262. Because of this improvement and to keep the model consistent with model 1 I decided to also remove CTRL2_2 and RISK2_2. It seems that these items measure the constructs in a different way than the others and removing them gives a much more accurate model.

	PTRST	UTRST2	CTRL2	RISK2	WILL2	SENS
PTRST_1	0,728					
PTRST_2	0,864					
PTRST_3	0,888					
PTRST_4	0,784					
UTRST2_1		0,866				
UTRST2_2		0,932				
UTRST2_3		0,952				
CTRL2_1			0,745			
CTRL2_2			0,775			
CTRL2_3			0,799			
CTRL2_4			0,75			
RISK2_1				0,807		
RISK2_2				0,783		
RISK2_3				0,865		
RISK2_4				0,707		
RISK2_5				0,749		
WILL2_1					0,932	
WILL2_2					0,966	
WILL2_3					0,912	
SENS_1						1

Table 3: Model 2 standardized factor loadings

Appendix G.2 shows the updated factor loadings after removing the two items. It's notable that after this change CTRL1_1 has a loading value of 0,609, which is lower than the 0,708 threshold. This item will be kept in the model, because it's recommended to use at least 3 indicators to give an accurate estimate (Marsh et al., 1998). The item is about who has control over who can get access to the data, while the other two control items are about how the data is used. It seems that people have more control over who has access to the data in scenario 2, compared to one, the mean value went from 2,973 to 5,281. The mean values of the other items also increased from scenario 1 to 2 (2,641 to 4,34 and 2,722 to 4,369). This can indicate that overall the respondents perceived much more control over who gets access to their data, and some more control over their data in general.

In addition to the factor loadings I looked at the modification indeces of the measurement models. They indicated that correlating some residual covariances could result in some improvement of model fit. I opted not to do this because I found little theoretical grounds to support correlating the items (Lei & Wu, 2007). The item pairs with a high modification index were also (mostly) different for model 1 and model 2. I argue that if the modifications can't be done for both models, I would be over fitting the model to the specific data.

4.2.2 Construct reliability and validity

The second step of assessing a reflective measurement model is assessing the internal consistency reliability (Hair et al., 2019). This refers to the: "the stability and consistency with which the instrument measures the concept" (Sekaran & Bougie, 2016, p.223). I do this by calculating the Cronbach's alpha and Composite reliability statistics for each constructs. The values are shown in table 4. Overall higher values mean higher reliability. Hair et al. (2019) claims that values from 0.7 to 0.9 range from satisfactory to good. The WILL constructs have values above 0.95. This can indicate that some items can be redundant, reducing construct validity

(Diamantopoulos et al., 2012). It can also indicate undesirable answer patterns, such as straight lining (Hair et al., 2019). I will test for construct validity later.

	Cronbach's alpha	Composite reliability	Average Variance Extracted
PTRST	0,885	0,895	0,684
UTRST1	0,956	0,958	0,883
CTRL1	0,907	0,910	0,768
RISK1	0,858	0,854	0,600
WILL1	0,964	0,965	0,901
UTRST2	0,939	0,943	0,844
CTRL2	0,805	0,853	0,649
RISK2	0,864	0,865	0,615
WILL2	0,955	0,956	0,878

Table 4: Reliability measures

This construct validity refers to the ”the extent to which the construct converges to explain the variance of its items” (Hair et al., 2019). I measure construct validity using the average variance extracted (AVE). The AVE values can also be found in table 4. All the constructs have an AVE of at least 0.5, this means that the constructs explain at least 50% of the variance in the items (Hair et al., 2019).

4.2.3 Discriminant validity

Discriminant validity is a requirement that is met if: ”a test not correlate too highly with measures from which it is supposed to differ” (Campbell, 1960). So if people would fill in the UTRST items very similar to the CTRL items, we would have low discriminant validity. The discriminant validity used to be assessed by comparing the AVE of each constructs to the regression loadings of the structural model (Fornell & Larcker, 1981). The AVE should be higher than the regression coefficient going into that construct. Comparing the AVE values from table 4 to the regression loadings in tables 8 and 9, I can confirm that the discriminant validity is met. A more recently developed and robust way of measuring discriminant validity is with the heterotrait-monotrait (HTMT) ratio (Henseler et al., 2015). A high HTMT ratio indicates that a construct measures is very similar to another construct. Henseler et al. (2015) recommends a cutoff value of 0.9. As is visible in tables 5 and 6 all the HTMT values are below 0.9.

	PTRST	UTRST1	CTRL1	RISK1	WILL1	SENS
PTRST	1					
UTRST1	0,473	1				
CTRL1	0,181	0,605	1			
RISK1	0,183	0,566	0,525	1		
WILL1	0,355	0,622	0,551	0,618	1	
SENS	NaN	NaN	NaN	NaN	NaN	NaN

Table 5: Model 1 HTMT

	PTRST	UTRST2	CTRL2	RISK2	WILL2	SENS
PTRST	1					
UTRST2	0,583	1				
CTRL2	0,473	0,801	1			
RISK2	0,379	0,63	0,753	1		
WILL2	0,462	0,699	0,803	0,706	1	
SENS	NaN	NaN	NaN	NaN	NaN	NaN

Table 6: Model 2 HTMT

4.3 Structural model

Now that we have a way of measuring the constructs in a reliable and valid, I want to take a look at the relationships between the constructs. To do this I use maximum likelihood algorithm in a structural equation model. The model I want to test for in this data is the conceptual model from chapter 2.4. The measurement models I use for this are the same as described in the previous chapter, so with a CTRL and RISK item removed. Appendix H shows the scripts made for RStudio that were used to generate these results.

4.3.1 Model fit

The first step of assessing the model is to look at overall model fit statistics. The standard way to do this is with the Chi-Square statistic (Anderson & Gerbing, 1988). There are also many comparative model fit measures available. I choose to follow the recommendation by Hu and Bentler (1999), which is to use the CFI, TLI and RMSEA, next to the Chi-Square statistic.

As discussed in chapter 4.1 some of my data is not exactly normally distributed. The maximum likelihood algorithm makes use of Pearson product-moment correlational techniques to treat ordinal data from my survey as if it was normally distributed continues data. If this data is not approximately normally distributed the chi square test, comparative fit indices and significance of parameter estimates will appear lower than they actually are (Finney & DiStefano, 2006). To compensate this effect Satorra and Bentler (2001) developed a scaling factor that takes this non normality into account. Table 7 shows all the model fit statistics. The ”robust” values are values with the Satorra-Bentler scaling applied. I will use the robust values to access the model fit.

Schreiber et al. (2006, p.330) made an overview of how structural equation model results are reported, which included a list of fit indices with cutoff values. I used these to access if my model fit was good enough. The Chi square/DF statistics are both < 3, which indicate a reasonable model fit. The comparative fit indices are very close to acceptable fit indeces (≥ 0.95 for CFI & TLI, ≤ 0.08 for RMSEA).

	Chi-square	Chi-Square/DF	CFI	TLI	RMSEA
Model 1	380,769	2,998	0,943	0,947	0,082
Model 1 robust	308,281	2,427	0,949	0,938	0,077
Model 2	446,728	3,517	0,922	0,906	0,093
Model 2 robust	301,613	2,374	0,936	0,923	0,083

Table 7: Model fit

4.3.2 Model 1

In the first model 5 out of 6 regression paths were found to be significant at the 0.1% level. Table 8 shows the coefficients of model 1. The coefficients also all have the expected sign. Control and trust in other users negatively influence risk, sensitivity of data positively influences risk and risk negatively influences the willingness to share data. Only trust in the platform did not influence risk in the way I expected it. I will go into more detail on this in chapter 4.3.4

		β	**p value**
H1	**CTRL1 ->RISK1**	-0,405	0,000
H2	**UTRST1 ->RISK1**	-0,47	0,000
H3	**PTRST ->RISK1**	0,014	0,827
H4	**PTRST ->CTRL1**	0,253	0,000
H5	**SENS ->RISK1**	0,263	0,000
H6	**RISK1 - >WILL1**	-0,704	0,000

Table 8: Model 1 results

4.3.3 Model 2

The second model has had very similar results. As can be seen in table 9. It supports the same 5 hypotheses as model 1, with H3 being rejected. There are slight differences in the coefficients of model 2 compared to one. Model 2 has a higher value for CTRL to RISK and lower values for UTRST and SENS to RISK. Also the impact of platform trust to control has increased. These results will be discussed in more detail in chapter 5.

		β	**P**
H1	**CTRL2 ->RISK2**	-0,515	0,000
H2	**UTRST2 ->RISK2**	-0,362	0,000
H3	**PTRST ->RISK2**	-0,016	0,824
H4	**PTRST ->CTRL2**	0,474	0,000
H5	**SENS ->RISK2**	0,196	0,000
H6	**RISK2 - >WILL2**	-0,747	0,000

Table 9: Model 2 results

4.3.4 Alternative models with PTRUST

In both models there was no significant relation between trust in the platform and perceived risk found. So I made a few alternative models to see if there was one that would better describe how platform trust influences the other constructs. The results of these alternative models performed on the data of model 1 can be found in table 10. Most of the results were similar for model 1 and 2, therefor only the result of model 1 is discussed here. Whenever there are notable differces between the models I will mention it. Appendix G.5 shows the results of the alternative models on the data of model 2.

The first modification I tried was removing the significant relationship between PTRST and CTRL. We've done this because we've had some doubts when creating the conceptual model about this relationship. There is theory supporting that people would feel like they had more control over the data if they trust the platform , but you can also argue that the feeling of control is separate from the trust in the platform. Because they both predict the perceived risk. So any impact that trust in the platform would have on control can be explained by perceiving more risk. So I tested the model with the relationship removed on my data. It proved to have a large increase in model fit, the Chi square/DF decreased from 2,427 to 1,820 in model 1. There were also some small effect on the factor loadings.

Because I was still questioning the relationship between PTRST and CTRL I also tried reversing this relationship. It was strange the model fit improved that much when removing a significant relationship. Because of this I tried reversing this relationship. This also resulted in an improvement in model fit compared to the initial model, but not as much as removing the relationship. So my data suggests that more feeling of control, would cause more trust in the platform.

After this I removed the insignificant relation of PTRST and RISK. This however made the model fit a lot worse (Chi square/DF 3,638). This would not be acceptable. Interestingly for the data of model 2 the model fit would not get worse, but stay similar to the original model. We've also created an alternative model with PTRST removed completely. This resulted in a substantial improvement in model fit for model 1 and a similar model fit for model 2.

Overall changing the way trust in the platform is used in the model did bring some notable improvements in model fit. In chapter 5 I will go into why this might be the case.

		Model 1		**PTRST ->CTRL removed**		**PTRST ->CTRL reversed**		**PTRST ->RISK removed**		**PTRST removed**	
	Chi square/DF	2,427		1,820		2,108		3,638		2,057	
			P		**P**		**P**		**P**		**P**
H1	**CTRL1 ->RISK1**	-0,405	0,000	-0,342	0,000	-0,344	0,000	-0,355	0,000	-0,343	0,000
H2	**UTRST1 ->RISK1**	-0,470	0,000	-0,434	0,000	-0,428	0,000	-0,406	0,000	-0,425	0,000
H3	**PTRST ->CTRL1**	0,253	0,000					0,192	0,005		
H3*	**CTRL1 ->PTRST**					0,240	0,000				
H4	**PTRST ->RISK1**	0,014	0,827	0,020	0,734	0,018	0,758				
H5	**SENS ->RISK1**	0,263	0,000	0,245	0,000	0,244	0,000	0,243	0,000	0,245	0,000
H6	**RISK1 - >WILL1**	-0,704	0,000	-0,694	0,000	-0,695	0,000	-0,676	0,000	-0,695	0,000

Table 10: Modifications to PTRST in model 1

4.3.5 Impact of demographic variables

Besides the variables in the conceptual model I asked the respondents for information about their experience, industry role, sector they supply to and familiarity with IoT and data marketplaces. In this chapter I will take a look into if any of these variables has an impact on the conceptual model. This is done by first checking for direct regression paths from the demographic variables to the latent variables. After this I take a closer look at some variables by filtering the dataset to only include a certain subset.

Table 11 shows the path coefficients of the three ordinal demographic variables (years of experience, familiarity with IoT and familiarity with data marketplaces). The variables and paths were added one at a time, so the effect on the model was measured in isolation.

	SENS	**PTRST**	**UTRST1**	**UTRST2**	**CTRL1**	**CTRL2**	**RISK1**	**RISK2**	**WILL1**	**WILL2**
Experience	0,095*	-	-	-	-	-	-	-0,087**	-	-
Familiarity IOT	0,113*	0,187***	0,182***	0,164**	0,138**	0,112*	-	-	0,077**	0,099**
Familiarity data marketplaces	-	0,229***	0,299***	0,236***	0,247***	0,144**	-	-	0,157***	0,125***

Table 11: Regression of demographic variables to latent constructs, - p$\geq$ 0.1, $*p<0.1, **p<0.05, ***p<0.01$

As we can see from table 11 the experience only has a minor impact on some of the variables. There is only some weak evidence that experience has a positive impact on sensitivity and moderate evidence that experience reduces the perceived risk. The familiarity had much more impact on the model. It seems that people that are familiar with IoT are more trusting to the

platform, trusting to other users and perceive more control. People that say they are familiar with data marketplaces are even more trusting and perceive even more control. The impact on the willingness to share data is also quite high for both the familiarity variables.
In addition to including the demographic variables to the model I also tested what it would look like if I filtered the data based. The results of this can be found in appendix I. I've tested 6 different groups:

1. High familiarity with data marketplaces
2. Low familiarity with IoT
3. Low sensitivity
4. High sensitivity
5. Middle management
6. Experienced respondents

For the high familiarity group I choose to select respondents that rated their familiarity with data marketplaces "Very familiar" or "Extremely familiar". This group was chosen separately because high familiarity seemed to increase the trust, control and willingness to share, as can be seen from table 11. Selecting only this group has left us with a sample size of only 78. Despite this it still resulted in some regression coefficients with very low P values. Some notable differences with this group compared to the complete sample are: An increase in correlation between PTRST to CTRL and from CTRL to RISK for the first model. A slight decrease in correlation of the same paths for model 2, and a decrease in correlation from RISK2 to WILL2. Interestingly it was harder to detect a correlation between SENS and RISK, especially for model 2.

The low familiarity group consisted of the responses with the familiarity in internet of things as "Not familiar at all" or "Slightly familiar", which was a sample of 109 responses. For the low familiarity group the most notable difference was that the sensitivity of data was a better predictor of risk, compared to the base models.

I also tested the group with low and high sensitivity seperately. With low having values 1 and 2 on the 5 point scale between not not sensitive and very sensitive, and high having a value of 4 and 5 on this scale. The low sensitivity group had a sample size that was too small get accurate results. The high sensitivity group had a sample size of 183, and gave some mixed results. For the first model the correlation between CTRL and RISK increased and the correlation between UTRST and RISK decreased, while for the second the correlation between CTRL and RISK decreased and the correlation between UTRST and RISK increased. This can indicate that in situations where the data is sensitive and there is less trust, perceived control becomes more important and in situations where there are more control mechanisms, perceived trust becomes more important.

To test if the industry role had an effect on the results I tested the group of only middle managers (N=142). This group would likely be most knowledgeable about how a data marketplace would actually work in a their company and have an advisory role in the adoption. The results show that middle managers value perceived control a bit less, compared to trust between other users. In addition to this I also tested what the most experienced group (5+ years of experience) would think. These are likely also older people with more experience using existing methods. This group answered very similarly to the base group, with an increase in the importance of sensitivity of data.

4.3.6 Mediation

In the conceptual model all the relationships the willingness to share are done through the risk factor. I found this to make the most theoretical sense, however what if some of the variables also have a direct effect to willingness to share? To test this I used mediation analysis (MacKinnon et al., 2007).

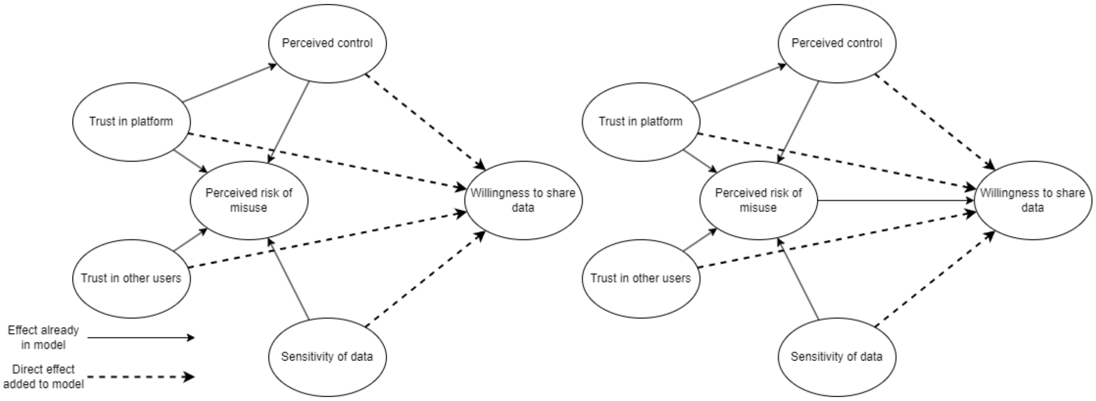


Figure 7: Mediation analysis (Left shows only direct effects, right shows direct and indirect effects)

The first step to this is detecting if there are direct effects, without the mediation variable (RISK) (Baron & Kenny, 1986), the left side of the figure 7 visualizes what this looks like. The first column of table 12 shows the results of this. It shows that there are significant regression paths from CTRL, UTRST and SENS to WILL in both models. In addition to this there is even a path significant at the 10% confidence interval from PTRST to WILL in the first model. This means that we can continue to step 2 where we check for the mediating effect of RISK.

It was already clear from previous results in 4.3 that there were significant paths from CTRL, UTRST and SENS to RISK and from RISK to WILL. In the mediating analysis I will add the direct paths to the existing model and check if these paths are still significant. The right side of figure 7 shows what this model looks like. If risk fully mediates these relationships, the direct relations to willingness to share, would have a coefficient of 0 (Baron & Kenny, 1986). The indirect effects consists of the effects from CTRL, UTRST, PTRST and SENS to RISK, multiplied by the effect from RISK to WILL. The total effects are the direct effects + the indirect effects.

	Direct effects mediation removed	**Direct effects including mediation**	**Indirect effects**	**Total effects**
CTRL1->(RISK1->)WILL1	0,309***	0,186***	0,122***	0,308
CTRL2->(RISK2->)WILL2	4,810***	0,323***	0,146***	0,469
UTRST1->(RISK1->)WILL1	0,450***	0,260***	0,160***	0,420
UTRST2->(RISK2->)WILL2	0,329***	0,252***	0,100***	0,360
PTRST->(RISK1->)WILL1	0,113*	0,136**	-	-
PTRST->(RISK2->)WILL2	-	-	-	-
SENS->(RISK1->)WILL1	-0,181***	-0,091*	-0,090**	-0,271
SENS->(RISK2->)WILL2	-0,175***	-0,119***	-0,056***	-0,175

Table 12: Mediation analysis, - p$\geq$ 0.1, $*p$<0.1, $**p$<0.05, $***p$<0.01

From table 12 we can see that there are significant paths from all the factors (with the exception of PTRST), both direct and indirect. This means that risk partially mediates the impact on willingness to share. For the perceived control, risk mediates about 1/3 of the total effects to willingness to share, 40% for model 1 and 31% for model 2. The trust in other users has 38% and 28% of it's total effects mediated through risk. Sensitivity of data has 33% and 32% of it's effect mediated through risk. This means that risk definitely helps explaining why companies are willing to share data, but just looking at control, trust in other users and sensitivity of data can also give a lot of information.

5 Discussion

In this chapter I will interpret the results from chapter 4. Reflecting on how they confirm or reject the hypotheses from chapter 2.4, how this impacts the existing knowledge about data marketplaces, what the practical implications for designing or participating in data marketplaces are and what limitations of the study are.

5.1 Relations between control, trust and risk

The perceived risk of sharing data is still an understudied in the contexts of data marketplaces where trust is expected to be low between the users. In my conceptual model I defined four constructs that (directly) impact the perceived risk of sharing data: Perceived control, Trust in other users, Trust in platform and Sensitivity of data. For three out of four I found significant correlations.

The first two hypotheses are about the impact of control and trust in other users on perceived risk. My results heavily support these hypotheses, which state that higher trust in other users and higher perceived control will decrease the perceived risk. The factor loadings of these two constructs were the highest of the four constructs impacting risk. Control had loadings of $\beta =$ -0,405, -0,515 for models 1 and 2, and trust in other users had loadings of $\beta =$ -0,470, -0,362 for models 1 and 2. Answering the second sub question I can say that control and trust in other users negatively impact the perceived risk of sharing data.

The sensitivity of data also proved to be important in predicting the perceived risks. Overall the sensitivity of the data was rated to be fairly high, even after I state that the respondents can remove critically sensitive data points. This is important for the validity of my results. As stated in the theory chapter, if the respondents don't think the data is particularly sensitive, they won't think there is a lot of risk in sharing the data. My results also show this. In both models there are significant relations between sensitivity and perceived risk. $\beta =$ 0,263, 0,196 for models 1 and 2 respectively. I expect that sector specific contextual factors play a large role in determining the sensitivity of the data. I recommend further research that more directly compares the differences in data sensitivity in relation to willingness of sharing data across different industries.

The results of trust in the platform was not in line with what I expected. The regression path between platform trust and risk was found insignificant and coefficients of practically 0. My data does not show this connection. It could be possible that the trust in platform just does not directly impact the risk of sharing data, but only indirectly through perceived control as a mediating factor. It could also be that the trust in other users, perceived control and sensitivity of data are just much better predictors of risk. A study with a larger sample size could reveal this. Interestingly the trust in platform was perceived quite high, much higher than trust in other users and perceived control. A possible explanation for this is in the position of the question in the survey and the wording of the items. I will discuss this further in the limitations (chapter 5.4).

I did find a significant relation between trust in the platform and perceived control, with $\beta =$ 0,263, 0.474. Removing this relationship from the model however resulted in a notable improvement of model fit. From my data we can conclude that there is an affect of trust in the platform on perceived control.

My study very clearly shows that more control and trust in other users results in less risk. This can be seen in the coefficients in both models, but also in the absolute values. The first scenario the respondents perceived a lot less control and had less trust in other users. This model also had much higher values for risk and lower values for willingness to share data. The scenario with more trust and more control resulted in much lower average values for risk and higher for willingness to share data. In the model with higher trust and control I also noticed a slightly higher coefficient of control to risk (and lower from trust in other users and sensitivity to risk). This can indicate that the perceived control and use of control mechanisms become more important once there is a baseline level of trust between the users. These results can be used to answer the third sub question: ”How do different control mechanisms influence the perceived risk of sharing data”. In a situation with more control mechanisms I found the perceived risk to be much lower, which in turn corresponded with more willingness to share data.

As is stated in the literature review in chapter 2, there seems to be a trade off between control and trust when it comes to perceived risk. Where a lack of trust can be compensated by adding in more control. My study does not confirm or disprove this. The two scenarios sketched in my questionnaire showed a low trust, low control scenario and a high trust, high control scenario. The results clearly showed there was less risk perceived in the high/high scenario. A future study might include a low trust, high control and a high control, low trust scenario. This could reveal that people don’t need trust if the control mechanisms work, or people don’t need control mechanisms if they trust the other user. The interaction between these two variables could use more attention.

There are many different factors that can determine the willingness to share data. In chapter 2.2 I found that perceived benefits and organizational readiness are very important for willingness to share data. For the manufacturing industry having a lot of data available is not common (Frank et al., 2019). In this study I wanted to focus on the impact of risk on the willingness to share. Therefor I gave the respondents some assumptions: They would have a lot of data available to share, the data is well organized, they would receive enough money for their data to consider sharing it, the data they have is compatible with the marketplace and the marketplace is very easy to use. With these assumptions in mind I found a significant relation with a high correlation (β =-0,704, -0,747) between perceived risk and willingness to share data. Meaning the respondents that perceived lower risk would more willing to share their data. Interestingly the mean values of willingness to share were quite high, 3.83/7 and 4,78/7 for models 1 and 2 respectively. So if all the assumptions mentioned can be realized, people would be willing to share data.

5.2 Theoretical implications

Research on data marketplaces is still relatively new, especially in the context of business to business data sharing. The adoption of business data marketplaces was identified as an important area for research (Abbas et al., 2021). This research contributes to this by focusing directly on one adoption factor: perceived risk, and indirectly in another factor: trust. Both these concepts have been mentioned several times as important drivers for data marketplace adoption (De Prieelle et al., 2022; Müller et al., 2020; Susanty et al., 2018). This research confirms that perceived risk indeed helps predict a company’s willingness to share data. It also gives insights on what impacts perceived risk. The data from this research also confirms that companies are willing to share data, if preconditions such as receiving a fair price are met. With this research we can now better understand that perceived risk lowers the willingness to share data.

We can also better understand how this risk can be lowered. Using control mechanisms and creating an environment in which the platform users trust each other. I would even say that the biggest contribution of this study to the existing knowledge is defining relationships between trust, control, sensitivity of data and risk. With this knowledge researchers can focus on how perceived control and trust can be increased, which will in turn lead to a decrease in risk. Researchers can also look into why users perceive risk. This research shows that the trust in other users, perceived control and sensitivity of data do a good job of predicting the perceived risk, but there can be more reasoning behind why a company would be reluctant to sharing data.

The research also shows that perceived control, trust in other users and sensitivity of data not only affect the risk, but also directly influence the willingness to share data. In fact we can see that risk mediates about 1/3 of the effects of these variables to willingness to share. Meaning perceived risk is important in predicting the willingness to share data, but having a high amount of trust and perceived control can also directly increase the willingness to share data. This is also an addition to the complicated theory that explains why companies would or would not share data in a marketplace.

So far there has been very little research on sharing business data in a marketplace in the manufacturing industry. Most research on this has been done in the context of supply chain platforms and sharing data directly between business partners. This research can be used as a basis for future studies on sharing industrial IoT data. It provides some theoretical background on which factors influence the adoption of platforms and what data can be shared in an industrial setting. It also provides some basic descriptions of an industrial data marketplace and what data can be shared using it.

Another contribution that this study makes is verifying some of the exploratory studies by doing quantitative research. Studies such as Agahari et al. (2022),Holler et al. (2019), Reimsbach-Kounatze (2021) and De Prieelle et al. (2022) show that factors such as trust in other users, perceived control and risk have an influence on the willingness to share data. This study confirms this, using quantitative data. The data from this research also shows there is some willingness to share data if preconditions are met and give insights into what these preconditions are.

5.3 Practical implications

Industrial data marketplaces are still in an early stage of development. There have been and are some platforms available, but none have reached a high user base. This has to do with many factors such as: unclear benefits, lack of data available by manufacturing companies, uncertainty about risk of losing a competitive advantage. This study can help developers of data marketplaces realize about the importance of control mechanisms. The study clearly shows that the scenario with user verification and more control mechanisms resulted in lower risk and higher willingness to share data. If the marketplace facilitator can create trust between the users and give providers control over their data it's much more likely companies are willing to share their data. In addition to this it shows that people with knowledge of industrial IoT and data marketplaces are much more likely to share their data. They also perceive a higher amount of control and trust.The study also shows that the specific mechanisms of user verification and smart contracts provide more trust between users and give more control do data providers.

For regulators this research has shown that they need to make it possible for companies to have control over their data. New regulations such as the data act (European Commission, 2022) should make it easier to share data between companies. In future regulations the European Union and other governing bodies could look into how their regulations can help increase trust when sharing data and how they could increase the perceived control over data.

5.4 Limitations

While the study resulted in some significant outcomes, there are some limitations that need to be considered. Firstly the respondents were given many assumptions before asking about their opinion on control, trust, risk and willingness to share. For example: receiving a fair price for the data, having enough data available and systems being inter operable with the marketplace. The results of this study are only valid if these assumptions are met. I think the effect of these assumptions on the relations between control, trust and risk will be very small. The risk of sharing data would be the same regardless of how much they would receive for it. Only the assumption of having a lot of data available could have influence over these variables. The data in the scenarios is hypothetical. They might think differently about sharing data if they actually had the data available. All the assumptions have a much larger effect on the measurement of willingness to share data. I can clearly show that risk has an effect on the willingness to share data, but to determine how big this is you would need to compare it to other factors such as perceived benefits, ease of use and readiness.

As with any quantitative study, the results are only as good as the sample. This sample has about 1/3 of its respondents from the US, 1/3 from the UK and 1/3 from the EU. The cultural differences between these regions can influence the results, maybe some countries are more reluctant to sharing with other companies that others. I analyzed results separately for the US and UK respondents to test this, but unfortunately the sample size was to low to do get significant results. It's difficult to judge exactly how well the respondents can judge the risks of sharing data. My sample had a lot of middle managers and where moderately familiar with industrial iot and data marketplaces. I think people in this sample are capable of judging this risk and understand how much control and trust they would have in the scenario's. Only including upper and middle managers with good understanding of industrial IoT would probably be the most accurate representation of people that would actually decide on adopting a data marketplace. This would not allow me to have a large enough sample.

The respondents were recruited using the Prolific platform, where they would receive a small payment for completing the results. This could have influenced the results. By having a monetary incentive people tend to complete the survey as fast as possible. I was able to eliminate extreme cases of this where the completion time was much lower and including comprehension checks to make sure they read the descriptions. To some extent this could still have an effect on the results. There is also a chance that the respondents lie about their industry role, in order to participate in more surveys. I don't consider this to be very high as this would not only make them eligible for this study, but also ineligible for studies with other industry roles.

The concepts in the survey are fairly complex for people that might be unfamiliar with data marketplaces. I've simplified some of the concepts to make them more comprehensible in the survey. Especially the architecture of the platform and the use of smart contracts was difficult to explain concisely. When explaining these concepts there is a trade off between going into detail and having a compact description. A description with more detail can explain a concept better, but will likely result in losing the respondents attention. The limited ability to explain the concepts to the respondents can have some effect in how they interpret the questions. A qualitative study could deal better with this issue.

Also the formulation of the question items themselves has an effect on the constructs I try to measure. For all constructs I've reformulated items from other studies. Constructs are all slightly different in the setting of an industrial data marketplace compared to the study they are taken from. I suspect this is especially the case for the trust in platform items. As is explained in chapter 4.3.4 it resulted in some insignificant results. The items are about honesty, ability to fulfill commitments, confidence and clarity. From my description I provide little reason for the respondents not to trust the platform. Perhaps items that go more into how the platform would store and handle their data would give different results. I expect that different wording of the items resulted in different values for the trust in platform construct. Testing the that would be more in line with the theory.

The order of the questions also can have an effect in the results. I've measured the trust in the platform after explaining who the facilitator is and how the platform is set up. On its own this might be alright, but I then link this trust to handling data in separate scenarios. It could be that the respondent would have a different level of trust in the platform after knowing how it would handle the data. Another point where I think the order of questions has an impact in the results is in handling the two scenarios. The low trust & control scenario is always presented before the high trust & control scenario. This would make it logical that the second scenario would have much lower risk than the first. This makes comparing between the two scenarios difficult. A seperate study that provides respondents with a single random scenario could result into different results, where the differences between scenario's can be measured in a more valid way.

6 Conclusion

How do control, trust and risk affect a company's decision to share data in a data marketplace? To answer this question I performed a literature review, which resulted in a conceptual model with a set of hypotheses. This model was then tested through a questionnaire designed for management in the manufacturing industry. The results showed that high levels of trust in other users and perceived control correlate very well with low levels of perceived risk. A low level of perceived risk in turn correlates well with a high willingness to share data. So enabling mechanisms to increase trust and control should indirectly increase the willingness to share data on a marketplace. An important caveat to this is that the data is not considered too sensitive, as the sensitivity of data is also an important predictor of perceived risk.

In my study I presented the users with two scenarios: One with low trust and control, and another with higher levels of trust and control. Apart from the correlations between control, trust, risk and willingness to share data being very similar, I noticed an improvement in all measured variables in the high trust and control example. Understanding how these improvements in trust and control and trust can be made is crucial for the development of data marketplaces.

This research has given more clear definitions of the relations between control, trust and risk. More importantly it has provided empirical data that confirms that control and trust affect the perceived risk. The slow adoption of data marketplaces is a complex issue with many challenges. The benefits of sharing the data need to be clear and companies need to be ready to share the data. This study contributed to this literature by focusing on the perceived risk by data providers. It shows that it's important to create trust between users of a data marketplace and give them many control mechanisms. This will lead to more willingness to share data.

Future research can look deeper into the factors control and trust. How can a platform increase trust between users, or what control mechanisms are most effective for increasing perceived control. Another avenue that can be explored is the perceived risk. We know now that trust, control and sensitivity can increase the perceived risk, but we don't know what else might contribute to this.

References

Aaltonen, A., Alaimo, C., & Kallinikos, J. (2021). The Making of Data Commodities: Data Analytics as an Embedded Process. *Journal of Management Information Systems*, *38*(2), 401–429. https://doi.org/10.1080/07421222.2021.1912928

Abbas, A. E., Agahari, W., van de Ven, M., Zuiderwijk, A., & de Reuver, M. (2021). Business Data Sharing through Data Marketplaces: A Systematic Literature Review. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*(7), 3321–3339. https://doi.org/10.3390/jtaer16070180

Agahari, W., & de Reuver, M. (2022). Rethinking consumers' data sharing decisions with the emergence of multi-party computation: an experimental design for evaluation. *European Conference on Information Systems*.

Agahari, W., Ofe, H., & de Reuver, M. (2022). It is not (only) about privacy: How multi-party computation redefines control, trust, and risk in data sharing. *Electronic Markets*. https://doi.org/10.1007/s12525-022-00572-w

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411–423. https://doi.org/10.1037/0033-2909.103.3.411

Andres, S., & Laoutaris, N. (2022). A Survey of Data Marketplaces and Their Business Models. *SIGMOD Record*, *51*(3), 18–29.

Azarm-Daigle, M., Kuziemsky, C., & Peyton, L. (2015). A Review of Cross Organizational Healthcare Data Sharing. *Procedia Computer Science*, *63*, 425–432. https://doi.org/10.1016/j.procs.2015.08.363

Bajoudah, S., Dong, C., & Missier, P. (2019). Toward a Decentralized, Trust-Less Marketplace for Brokered IoT Data Trading Using Blockchain. *2019 IEEE International Conference on Blockchain (Blockchain)*, 339–346. https://doi.org/10.1109/Blockchain.2019.00053

Baldwin, C. Y., & Woodard, C. J. (2009). The Architecture of Platforms: A Unified View. In *Platforms, markets and innovation* (pp. 19–44). Edward Elgar Publishing. https://doi.org/10.4337/9781849803311.00008

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Beldad, A., de Jong, M., & Steehouder, M. (2010). How shall I trust the faceless and the intangible? A literature review on the antecedents of online trust. *Computers in Human Behavior*, *26*(5), 857–869. https://doi.org/10.1016/j.chb.2010.03.013

Beverungen, D., Hess, T., Köster, A., & Lehrer, C. (2022). From private digital platforms to public data spaces: implications for the digital transformation. *Electronic Markets*, *32*(2), 493–501. https://doi.org/10.1007/s12525-022-00553-z

Boudreau, K. (2010). Open Platform Strategies and Innovation: Granting Access vs. Devolving Control. *Management Science*, *56*(10), 1849–1872. https://doi.org/10.1287/mnsc.1100.1215

Broekhuizen, T., Emrich, O., Gijsenberg, M., Broekhuis, M., Donkers, B., & Sloot, L. (2021). Digital platform openness: Drivers, dimensions and outcomes. *Journal of Business Research*, *122*, 902–914. https://doi.org/10.1016/j.jbusres.2019.07.001

Brown, T. A. (2015). Introduction to CFA. In *Confirmatory factor analysis for applied research* (pp. 35–48). Guilford publications.

Calancea, C. G., & Alboaie, L. (2021). Techniques to Improve B2B Data Governance Using FAIR Principles. *Mathematics*, *9*(9), 1059. https://doi.org/10.3390/math9091059

Campbell, D. T. (1960). Recommendations for APA test standards regarding construct, trait, or discriminant validity. *American Psychologist*, *15*(8), 546–553. https://doi.org/10.1037/h0048255

Cárdenas, E. R., & Molano, V. M. (2022). Business Registration Data as the Best Vehicle to Achieve KYC and AML for Business. https://doi.org/10.1007/978-3-030-85817-9{_}13

Chang, Y.-Y., Lin, S.-C., Yen, D. C., & Hung, J.-W. (2020). The trust model of enterprise purchasing for B2B e-marketplaces. *Computer Standards & Interfaces*, *70*, 103422. https://doi.org/10.1016/j.csi.2020.103422

Chen, Y.-H., Lin, T.-P., & Yen, D. C. (2014). How to facilitate inter-organizational knowledge sharing: The impact of trust. *Information & Management*, *51*(5), 568–578. https://doi.org/10.1016/j.im.2014.03.007

Cheng, J.-H., Chen, S.-W., & Chen, F.-Y. (2013). Exploring how inter-organizational relational benefits affect information sharing in supply chains. *Information Technology and Management*, *14*(4), 283–294. https://doi.org/10.1007/s10799-013-0165-x

Christidis, J., Karkazis, P. A., Papadopoulos, P., & Leligou, H. C. ((2022). Decentralized Blockchain-Based IoT Data Marketplaces. *Journal of Sensor and Actuator Networks*, *11*(3), 39. https://doi.org/10.3390/jsan11030039

Chwelos, P., Benbasat, I., & Dexter, A. S. (2001). Research Report: Empirical Test of an EDI Adoption Model. *Information Systems Research*, *12*(3), 304–321. https://doi.org/10.1287/isre.12.3.304.9708

Creswell, J. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications.

Dai, H.-N., Wang, H., Xu, G., Wan, J., & Imran, M. (2020). Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies. *Enterprise Information Systems*, *14*(9-10), 1279–1303. https://doi.org/10.1080/17517575.2019.1633689

De Prieelle, F., de Reuver, M., & Rezaei, J. (2022). The Role of Ecosystem Data Governance in Adoption of Data Platforms by Internet-of-Things Data Providers: Case of Dutch Horticulture Industry. *IEEE Transactions on Engineering Management*, *69*(4), 940–950. https://doi.org/10.1109/TEM.2020.2966024

De Reuver, M., Sørensen, C., & Basole, R. C. (2018). The Digital Platform: A Research Agenda. *Journal of Information Technology*, *33*(2), 124–135. https://doi.org/10.1057/s41265-016-0033-3

Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. *Journal of the Academy of Marketing Science*, *40*(3), 434–449. https://doi.org/10.1007/s11747-011-0300-3

Dwaikat, N. Y., Money, A. H., Behashti, H. M., & Salehi-Sangari, E. (2018). How does information sharing affect first-tier suppliers' flexibility? Evidence from the automotive industry in Sweden. *Production Planning & Control*, *29*(4), 289–300. https://doi.org/10.1080/09537287.2017.1420261

Eisenmann, T., Parker, G., & Van Alstyne, M. W. (2006). Strategies for two-sided markets. *Harvard business review*, *84*(10), 92.

Eisenmann, T. R., Parker, G., & Van Alstyne, M. (2009). Opening Platforms: How, When and Why? In *Platforms, markets and innovation* (pp. 131–153). Edward Elgar Publishing. https://doi.org/10.4337/9781849803311.00013

European Commission. (2016, April 27). *General data protection regulation*. Retrieved October 11, 2022, from https://eur-lex.europa.eu/eli/reg/2016/679/oj

European Commission. (2022, February 23). *Data act: Commission proposes measures for a fair and innovative data economy*. Retrieved May 8, 2023, from https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113

Figueredo, K., Seed, D., & Wang, C. (2022). A Scalable, Standards-Based Approach for IoT Data Sharing and Ecosystem Monetization. *IEEE Internet of Things Journal*, *9*(8), 5645–5652. https://doi.org/10.1109/JIOT.2020.3023035

Finney, J., Sara, & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In *Structural equation modeling: A second course* (pp. 269–314).

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *18*(1), 39–50. https://doi.org/10.1177/002224378101800104

Frank, A. G., Dalenogare, L. S., & Ayala, N. F. (2019). Industry 4.0 technologies: Implementation patterns in manufacturing companies. *International Journal of Production Economics*, *210*, 15–26. https://doi.org/10.1016/j.ijpe.2019.01.004

Fu, H.-P., Chang, T.-H., Ku, C.-Y., Chang, T.-S., & Huang, C.-H. (2014). The critical success factors affecting the adoption of inter-organization systems by SMEs. *Journal of Business & Industrial Marketing*, *29*(5), 400–416. https://doi.org/10.1108/JBIM-04-2012-0070

Gawer, A. (2014). Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research Policy*, *43*(7), 1239–1249. https://doi.org/10.1016/j.respol.2014.03.006

González, V., Sánchez, L., Lanza, J., Santana, J. R., Sotres, P., & García, A. E. (2023). On the use of Blockchain to enable a highly scalable Internet of Things Data Marketplace. *Internet of Things*, *22*, 100722. https://doi.org/10.1016/j.iot.2023.100722

Gupta, P., Dedeoglu, V., Kanhere, S. S., & Jurdak, R. (2021). Towards a blockchain powered IoT data marketplace. *2021 International Conference on COMmunication Systems & NETworkS (COMSNETS)*, 366–368. https://doi.org/10.1109/COMSNETS51098.2021.9352865

Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2021). An Introduction to Structural Equation Modeling. https://doi.org/10.1007/978-3-030-80519-7{_}1

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, *31*(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Holler, M., Vogt, H., & Barth, L. (2019). Exploring the Willingness-to-Share Data of Digitized Products in B2B Manufacturing Industries. *Humanizing Technology for a Sustainable Society*, 1065–1072. https://doi.org/10.18690/978-961-286-280-0.56

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jovanovic, M., Sjödin, D., & Parida, V. (2021). Co-evolution of platform architecture, platform services, and platform governance: Expanding the platform value of industrial digital platforms. *Technovation*, 102218. https://doi.org/10.1016/j.technovation.2020.102218

Karhu, K., Gustafsson, R., & Lyytinen, K. (2018). Exploiting and Defending Open Digital Platforms with Boundary Resources: Android's Five Platform Forks. *Information Systems Research*, *29*(2), 479–497. https://doi.org/10.1287/isre.2018.0786

Kazantsev, N., Pishchulov, G., Mehandjiev, N., Sampaio, P., & Zolkiewski, J. (2022). Investigating barriers to demand-driven SME collaboration in low-volume high-variability manufacturing. *Supply Chain Management: An International Journal*, *27*(2), 265–282. https://doi.org/10.1108/SCM-10-2021-0486

Kehr, F., Kowatsch, T., Wentzel, D., & Fleisch, E. (2015). Blissfully ignorant: the effects of general privacy concerns, general institutional trust, and affect in the privacy calculus. *Information Systems Journal*, *25*(6), 607–635. https://doi.org/10.1111/isj.12062

Kerber, W. (2018). Data governance in connected cars: the problem of access to in-vehicle data. *J. Intell. Prop. Info. Tech. & Elec. Com. L.*, *9*, 310.

Koutroumpis, P., Leiponen, A., & Thomas, L. D. W. (2020). Markets for data. *Industrial and Corporate Change*, *29*(3), 645–660. https://doi.org/10.1093/icc/dtaa002

Kuan, K. K., & Chau, P. Y. (2001). A perception-based model for EDI adoption in small businesses using a technology–organization–environment framework. *Information & Management*, *38*(8), 507–521. https://doi.org/10.1016/S0378-7206(01)00073-8

Lade, P., Ghosh, R., & Srinivasan, S. (2017). Manufacturing Analytics and Industrial Internet of Things. *IEEE Intelligent Systems*, *32*(3), 74–79. https://doi.org/10.1109/MIS.2017.49

Lasi, H., Fettke, P., Kemper, H.-G., Feld, T., & Hoffmann, M. (2014). Industry 4.0. *Business & Information Systems Engineering*, *6*(4), 239–242. https://doi.org/10.1007/s12599-014-0334-4

Lee, H. L., So, K. C., & Tang, C. S. (2000). The Value of Information Sharing in a Two-Level Supply Chain. *Management Science*, *46*(5), 626–643. https://doi.org/10.1287/mnsc.46.5.626.12047

Lei, P.-W., & Wu, Q. (2007). Introduction to Structural Equation Modeling: Issues and Practical Considerations. *Educational Measurement: Issues and Practice*, *26*(3), 33–43. https://doi.org/10.1111/j.1745-3992.2007.00099.x

Li, J., Sikora, R., Shaw, M. J., & Woo Tan, G. (2006). A strategic analysis of inter organizational information sharing. *Decision Support Systems*, *42*(1), 251–266. https://doi.org/10.1016/j.dss.2004.12.003

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation Analysis. *Annual Review of Psychology*, *58*(1), 593–614. https://doi.org/10.1146/annurev.psych.58.110405.085542

Mao, W., Zheng, Z., & Wu, F. (2019). Pricing for Revenue Maximization in IoT Data Markets: An Information Design Perspective. *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, 1837–1845. https://doi.org/10.1109/INFOCOM.2019.8737571

Marsh, H. W., Hau, K.-T., Balla, J. R., & Grayson, D. (1998). Is More Ever Too Much? The Number of Indicators per Factor in Confirmatory Factor Analysis. *Multivariate Behavioral Research*, *33*(2), 181–220. https://doi.org/10.1207/s15327906mbr3302{_}1

Martin, K. D., Borah, A., & Palmatier, R. W. (2017). Data Privacy: Effects on Customer and Firm Performance. *Journal of Marketing*, *81*(1), 36–58. https://doi.org/10.1509/jm.15.0497

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model Of Organizational Trust. *Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.5465/amr.1995.9508080335

Misura, K., & Zagar, M. (2016). Data marketplace for Internet of Things. *2016 International Conference on Smart Systems and Technologies (SST)*, 255–260. https://doi.org/10.1109/SST.2016.7765669

More, S., & Alber, L. (2022). YOU SHALL NOT COMPUTE on my Data: Access Policies for Privacy-Preserving Data Marketplaces and an Implementation for a Distributed Market using MPC. *Proceedings of the 17th International Conference on Availability, Reliability and Security*, 1–8. https://doi.org/10.1145/3538969.3544445

Mosterd, L., Sobota, V. C., van de Kaa, G., Ding, A. Y., & de Reuver, M. (2021). Context dependent trade-offs around platform-to-platform openness: The case of the Internet of Things. *Technovation*, *108*, 102331. https://doi.org/10.1016/j.technovation.2021.102331

Mouzakitis, S., & Askounis, D. (2010). A Knowledge-Based Framework for Measuring Organizational Readiness for the Adoption of B2B Integration Systems. *Information Systems Management*, *27*(3), 253–266. https://doi.org/10.1080/10580530.2010.493842

Müller, J. M., Veile, J. W., & Voigt, K.-I. (2020). Prerequisites and incentives for digital information sharing in Industry 4.0 – An international comparison across data types. *Computers & Industrial Engineering*, *148*. https://doi.org/10.1016/j.cie.2020.106733

Narayanan, A., & Shmatikov, V. (2008). Robust De-anonymization of Large Sparse Datasets. *2008 IEEE Symposium on Security and Privacy (sp 2008)*, 111–125. https://doi.org/10.1109/SP.2008.33

Nazifa, T. H., & Ramachandran, K. (2019). Information Sharing in Supply Chain Management: A Case Study Between the Cooperative Partners in Manufacturing Industry. *Journal* *of System and Management Sciences*, *9*(1), 19–47. https://doi.org/10.33168/JSMS.2019.0102

Noorian, Z., Iyilade, J., Mohkami, M., & Vassileva, J. (2014). Trust Mechanism for Enforcing Compliance to Secondary Data Use Contracts. *2014 IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications*, 519–526. https://doi.org/10.1109/TrustCom.2014.66

Otto, B., Steinbuß, S., Teuscher, A., & Lohmann, S. (2019). *Reference architecture model* (tech. rep.). International data spaces association.

Penttinen, E., Halme, M., Lyytinen, K., & Myllynen, N. (2018). What Influences Choice of Business-to-Business Connectivity Platforms? *International Journal of Electronic Commerce*, *22*(4), 479–509. https://doi.org/10.1080/10864415.2018.1485083

Ramadhan, F., & Samadhi, T. M. A. A. (2016). Inter-organizational trust and knowledge sharing model between manufacturer and supplier in the automotive industry. *2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 856–860. https://doi.org/10.1109/IEEM.2016.7797998

Reimsbach-Kounatze, C. (2021). Enhancing access to and sharing of data: Striking the balance between openness and control over data. In *Data access, consumer interests and public welfare* (pp. 25–68). Nomos Verlagsgesellschaft mbH & Co. KG. https://doi.org/10.5771/9783748924999-25

Richter, H., & Slowinski, P. R. (2019). The Data Sharing Economy: On the Emergence of New Intermediaries. *IIC - International Review of Intellectual Property and Competition Law*, *50*(1), 4–29. https://doi.org/10.1007/s40319-018-00777-7

Rochet, J.-C., & Tirole, J. (2003). Platform Competition in Two-Sided Markets. *Journal of the European Economic Association*, *1*(4), 990–1029. https://doi.org/10.1162/154247603322493212

Roman, D., & Stefano, G. (2016). Towards a Reference Architecture for Trusted Data Marketplaces: The Credit Scoring Perspective. *2016 2nd International Conference on Open and Big Data (OBD)*, 95–101. https://doi.org/10.1109/OBD.2016.21

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*(4), 507–514. https://doi.org/10.1007/BF02296192

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *The Journal of Educational Research*, *99*(6), 323–338. https://doi.org/10.3200/JOER.99.6.323-338

Sekaran, U., & Bougie, R. (2016). *Research methods for business: A skill building approach*. John Wiley & Sons.

Spiekermann, M. (2019). Data Marketplaces: Trends and Monetisation of Data Goods. *Intereconomics*, *54*(4), 208–216. https://doi.org/10.1007/s10272-019-0826-z

Susanty, A., Sirait, N. M., & Bakhtiar, A. (2018). The relationship between information sharing, informal contracts and trust on performance of supply chain management in the SMEs of batik. *Measuring Business Excellence*, *22*(3), 292–314. https://doi.org/10.1108/MBE-05-2017-0019

Tadelis, S. (2016). Reputation and Feedback Systems in Online Platform Markets. *Annual Review of Economics*, *8*(1), 321–340. https://doi.org/10.1146/annurev-economics-080315-015325

Tiwana, A. (2014). Platform Governance. In *Platform ecosystems* (pp. 117–151). Elsevier. https://doi.org/10.1016/B978-0-12-408066-9.00006-0

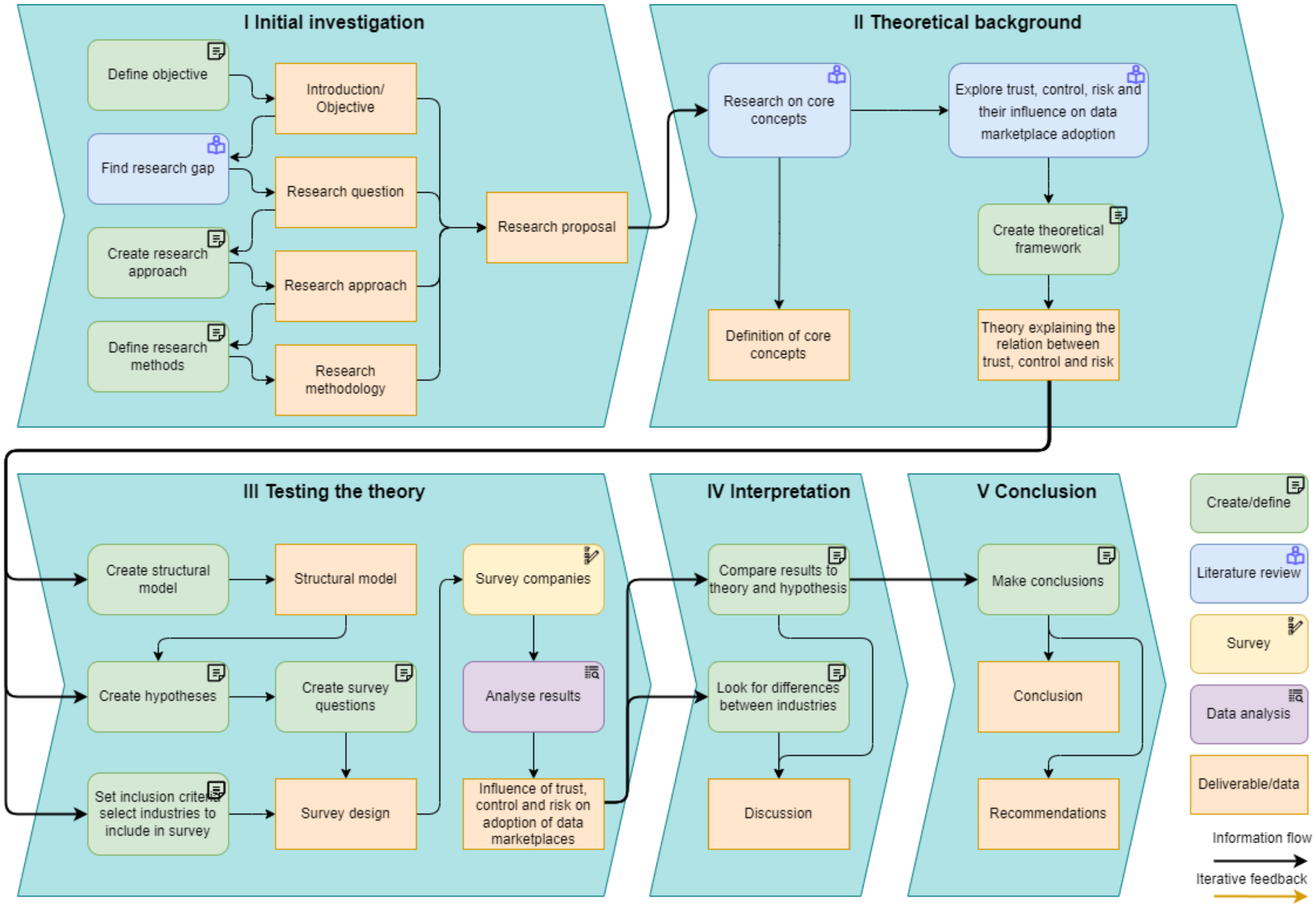
Tiwana, A., Konsynski, B., & Bush, A. A. (2010). Platform Evolution: Coevolution of Platform Architecture, Governance, and Environmental Dynamics. *Information Systems Research*, *21*(4), 675–687. https://doi.org/10.1287/isre.1100.0323

Tornatzky, L. G., & Fleischer, M. (1990). *The Processes of Technological Innovation*. Lexington Books.

Weber, I., Xu, X., Riveret, R., Governatori, G., Ponomarev, A., & Mendling, J. (2016). Untrusted Business Process Monitoring and Execution Using Blockchain. https://doi.org/10.1007/978-3-319-45348-4{_}19

Xu, H., Dinev, T., Smith, J., & Hart, P. (2008). Examining the Formation of Individual's Privacy Concerns: Toward an Integrative View. *ICIS*.

A Research Flow Diagram



B Survey descriptions

B.1 Opening statement

The impact of control, trust and risk on business participation in data marketplaces
You are being invited to participate in a research study titled: “The Impact of control, trust and risk on business participation in data marketplaces”. This study is being done by Floris Kool and Mark de Reuver from the TU Delft.

The purpose of this research study is to create a better understanding of why manufacturing companies would or would not participate in industrial IoT data marketplaces, and will take you approximately 15 minutes to complete. We will be asking you to provide some information about your company, your familiarity with the concepts of industrial IoT and Data marketplaces, and your opinion on sharing data through a data marketplace.

The data will be used for writing a master thesis, which will be published on the TU Delft repository. The data from the survey will be made publicly accessible on the 4TU.ResearchData repository. Free text questions will not be made publicly accessible.

The questionnaire is designed to be anonymous. We will not collect any information that we will be able to identify you with.

Your participation in this study is entirely voluntary and you can withdraw at any time. You are free to omit any questions.

By selecting “Yes” to the question below and completing the full questionnaire, you will accept that we process your data as described.

Contact details corresponding researcher:
Floris Kool
F.S.R.Kool@student.tudelft.nl

Contact details responsible researcher:
Mark de Reuver
G.A.deReuver@tudelft.nl

B.2 Data marketplace description

Sharing your company’s data in a data marketplace
Imagine the company you work for has spent the past 5-10 years investing into smart manufacturing systems. All the applicable machines have sensors that collect data and all the steps in the manufacturing process get tracked and stored. You’ve also spent a lot of effort in organizing the data, so the systems are interoperable and you can access everything easily. In this hypothetical scenario you can simply log into a server and see the real time status of all your machines and active manufacturing processes, as well as all historical data. Any data you can think of that can be useful to track in a manufacturing process is available.

This data may be very valuable to you, but it can also create value for other (unrelated) companies. For example a company that uses a similar machine to create different products might be interested in seeing how you optimized it’s processes. Or a company that’s proficient in the use of data analytics can use your data to make their business processes more efficient. Because of this value, you could sell your data. This is what a data marketplace is for.

A data marketplace connects buyers (consumers) and sellers (providers). We are interested in what risks you perceive when sharing data in a marketplace and how data marketplaces could be designed so you trust the other party. Pricing might be important, but is not part of this questionnaire. You can assume that you will receive a fair price for your shared data. This means enough to cover any expenses that you need to make to share the data and some profit. Enough to consider sharing the data you're already producing. You can also assume that your company is ready to share the data. All systems are compatible and it takes little to no effort to train personnel to use the marketplace. It's really almost as simple as clicking a ”Share data” button.

B.3 Data marketplace facilitator and architecture description

Data marketplace facilitator
Before showing how you can share data, we think it's important to get to know the owner of the marketplace and how the marketplace is structured.

The party that's facilitating all the data exchange is a new scale up company. The company is independent from big tech and other companies and is purely interested in facilitating data exchange, making revenue with every transaction on the marketplace. This facilitating company has various geographically distributed locations to reach companies in many areas.

The facilitating company makes sure every transaction between the users gets stored on a blockchain. You can view this as a long list of transactions where every new transaction gets added to the top of the list. Every new addition to the list gets verified through a large network of computers. Once a facilitator has sent the transaction and it's been verified, it's permanently stored on the blockchain. There is no way to hack or change transactions stored on the blockchain.

To ensure no privacy sensitive data is traded regulators have access to this blockchain. These are government bodies in various countries. They audit and monitor if the facilitators are complying to local privacy laws.

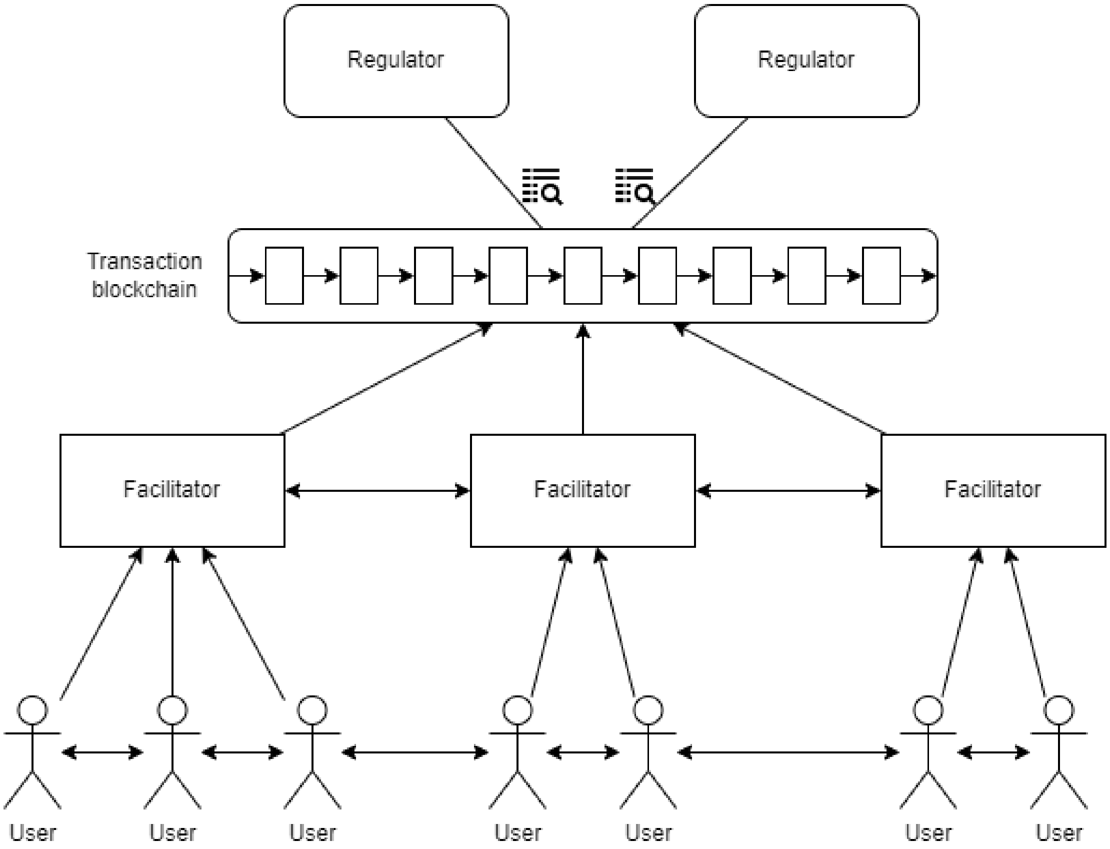


Figure 8: Data marketplace architecture. Image and description are adopted and simplified from Gupta et al. (2021)

Gupta, P., Dedeoglu, V., Kanhere, S. S., Jurdak, R. (2021). Towards a blockchain powered IoT data marketplace. 2021 International Conference on COMmunication Systems NETworkS (COMSNETS), 366–368.

B.4 Data sensitivity

Data sensitivity
Think about all the data you could generate in your manufacturing process. This includes all live and historical sensor data and some data that you use to keep track of your manufacturing process, this may include inventory data or live status of the manufacturing process. This data might be sensitive. It could give away an advantage to a competitor or contain personal information. Let's say you can remove some data points you consider very sensitive, for example the names of products or the inventory dataset.

B.5 Scenario 1

Scenario 1/2
In this data marketplace **everyone can create an account**. They would just need to enter their email address and basic profile in formation. You can see other companies profile, but there's no way of identifying who is actually purchasing your data.

The marketplace will allow you to select which data you'd like to share and for what price. You will also set the terms of use, where you state for example for what purpose the data may be used. This is legally binding, but you can't really know if the other users are complying. The data will then be listed on the marketplace.

Anyone who pays the price and agree to your terms of use will **download** your uploaded dataset.

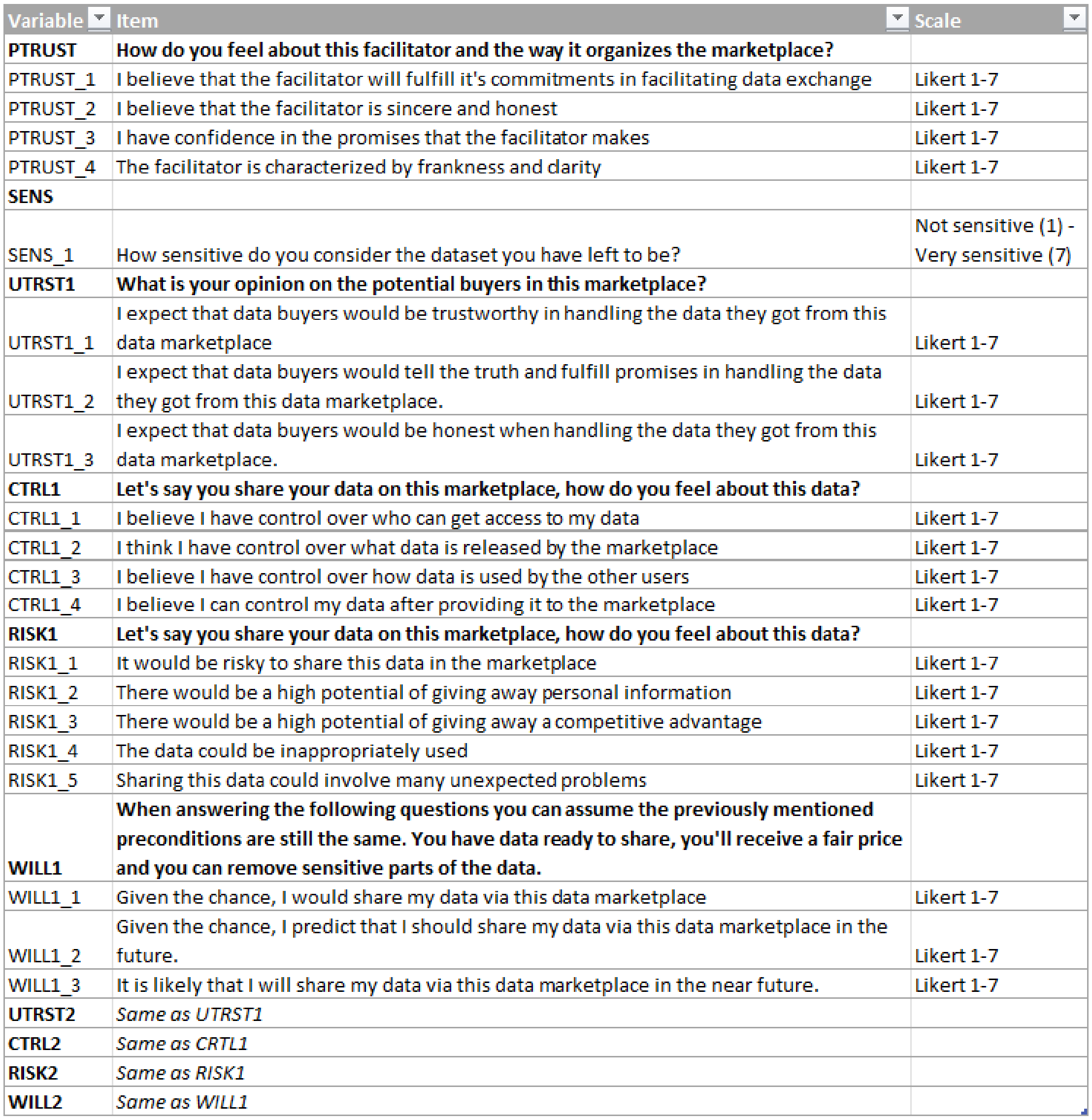
B.6 Scenario 2

Scenario 2/2
In this data marketplace users need to go through a textbfverification process. The process verifies the identity of the users and the company they work for. On this marketplace you can be fairly sure that you know who you're selling the data to.

The marketplace will allow you to select which data you'd like to share. You will also write down some terms of use, where you state for example for what purpose the data may be used, along with a suggested price. If someone wants to access your data they will have to send a **request**. In this request they have to state what data they want access to, what they intent to use your data for and what price they are willing to pay. You can textbfnegotiate these terms with the potential buyer until you reach an agreement.

Once you both agree on the data being sold, the terms of use and the price, a smart contract will be created. This is a self-executing contract. It can for example grant **access to your data** for a set amount of time or allow for automatic payment every time the user accesses the data.

C Survey items



Variable	Item	Scale
PTRUST	**How do you feel about this facilitator and the way it organizes the marketplace?**	
PTRUST_1	I believe that the facilitator will fulfill it's commitments in facilitating data exchange	Likert 1-7
PTRUST_2	I believe that the facilitator is sincere and honest	Likert 1-7
PTRUST_3	I have confidence in the promises that the facilitator makes	Likert 1-7
PTRUST_4	The facilitator is characterized by frankness and darity	Likert 1-7
SENS		
SENS_1	How sensitive do you consider the dataset you have left to be?	Not sensitive (1) - Very sensitive (7)
UTRST1	**What is your opinion on the potential buyers in this marketplace?**	
UTRST1_1	I expect that data buyers would be trustworthy in handling the data they got from this data marketplace	Likert 1-7
UTRST1_2	I expect that data buyers would tell the truth and fulfill promises in handling the data they got from this data marketplace.	Likert 1-7
UTRST1_3	I expect that data buyers would be honest when handling the data they got from this data marketplace.	Likert 1-7
CTRL1	**Let's say you share your data on this marketplace, how do you feel about this data?**	
CTRL1_1	I believe I have control over who can get access to my data	Likert 1-7
CTRL1_2	I think I have control over what data is released by the marketplace	Likert 1-7
CTRL1_3	I believe I have control over how data is used by the other users	Likert 1-7
CTRL1_4	I believe I can control my data after providing it to the marketplace	Likert 1-7
RISK1	**Let's say you share your data on this marketplace, how do you feel about this data?**	
RISK1_1	It would be risky to share this data in the marketplace	Likert 1-7
RISK1_2	There would be a high potential of giving away personal information	Likert 1-7
RISK1_3	There would be a high potential of giving away a competitive advantage	Likert 1-7
RISK1_4	The data could be inappropriately used	Likert 1-7
RISK1_5	Sharing this data could involve many unexpected problems	Likert 1-7
WILL1	**When answering the following questions you can assume the previously mentioned preconditions are still the same. You have data ready to share, you'll receive a fair price and you can remove sensitive parts of the data.**	
WILL1_1	Given the chance, I would share my data via this data marketplace	Likert 1-7
WILL1_2	Given the chance, I predict that I should share my data via this data marketplace in the future.	Likert 1-7
WILL1_3	It is likely that I will share my data via this data marketplace in the near future.	Likert 1-7
UTRST2	*Same as UTRST1*	
CTRL2	*Same as CRTL1*	
RISK2	*Same as RISK1*	
WILL2	*Same as WILL1*	

D Demographic information

N = 295			
Variable	**Answer**	**Frequency**	**Percentage**
Country of residence	UK	127	43%
	USA	86	29%
	EU	82	28%
Industry role	Upper management	25	8%
	Middle management	142	48%
	Junior management	80	27%
	Consultant	12	4%
	Researcher	9	3%
	Other	27	9%
Years of experience	<1	26	9%
	1-2	55	19%
	3-5	77	26%
	6-10	62	21%
	10+	75	25%
Familiarity with IIoT	Not familiar at all	32	11%
	Slightly familiar	77	26%
	Moderately familiar	89	30%
	Very familiar	82	28%
	Extremely familiar	15	5%
Familiarity with data marketplaces	Not familiar at all	43	15%
	Slightly familiar	75	25%
	Moderately familiar	99	34%
	Very familiar	62	21%
	Extremely familiar	16	5%
Sector	Blank	10	3%
	Other	79	27%
	Manufacturing/B2B	118	40%
	Food	17	6%
	Medical	11	4%
	Consumers/Retail	37	13%
	Automotive	23	8%

E Descriptive statistics

	N	Mean	Std deviation	Skewness statistic	Kurtosis statistic
PTRST_1	295	5,4	0,949	-0,972	1,927
PTRST_2	295	5,037	1,150	-0,641	0,443
PTRST_3	295	5,064	1,226	-1,049	1,418
PTRST_4	295	4,908	1,155	-0,433	0,365
SENS_1	295	4,702	1,372	-0,451	-0,217
UTRST1_1	295	2,875	1,521	-0,014	-0,790
UTRST1_2	295	2,875	1,619	0,011	-0,852
UTRST1_3	295	2,892	1,583	-0,033	-0,861
CTRL1_1	295	2,973	1,738	0,649	-0,739
CTRL1_2	295	4,044	1,738	-0,268	-1,050
CTRL1_3	295	2,641	1,702	0,911	-0,169
CTRL1_4	295	2,722	1,628	0,856	-0,181
RISK1_1	295	4,881	1,359	-0,406	-0,269
RISK1_2	295	4,295	1,399	-0,147	-0,372
RISK1_3	295	5,075	1,381	-0,666	0,317
RISK1_4	295	5,319	1,343	-0,767	0,597
RISK1_5	295	5,064	1,314	-0,599	0,222
WILL1_1	295	3,915	1,557	-0,125	-0,890
WILL1_2	295	3,827	1,621	-0,027	-0,944
WILL1_3	295	3,759	1,616	0,030	-0,911
UTRST2_1	295	4,122	1,285	-0,878	0,700
UTRST2_2	294	3,983	1,359	-0,717	0,171
UTRST2_3	295	3,993	1,353	-0,809	0,359
CTRL2_1	295	5,281	1,345	-1,006	0,651
CTRL2_2	295	5,275	1,292	-1,095	1,113
CTRL2_3	294	4,34	1,663	-0,417	-0,819
CTRL2_4	295	4,369	1,632	-0,340	-0,757
RISK2_1	295	3,983	1,444	0,112	-0,486
RISK2_2	295	3,685	1,392	0,159	-0,551
RISK2_3	295	4,146	1,444	-0,080	-0,495
RISK2_4	295	4,342	1,441	-0,320	-0,320
RISK2_5	295	4,166	1,439	-0,066	-0,436
WILL2_1	295	4,786	1,423	-0,744	0,081
WILL2_2	295	4,803	1,460	-0,658	-0,171
WILL2_3	295	4,736	1,482	-0,542	-0,312

F Model 1 results

F.1 Factor loadings

	PTRST	UTRST1	CTRL1	RISK1	WILL1	SENS
PTRST_1	0,722					
PTRST_2	0,86					
PTRST_3	0,897					
PTRST_4	0,776					
UTRST1_1		0,914				
UTRST1_2		0,952				
UTRST1_3		0,948				
CTRL1_1			0,812			
CTRL1_2			0,507			
CTRL1_3			0,915			
CTRL1_4			0,901			
RISK1_1				0,831		
RISK1_2				0,626		
RISK1_3				0,751		
RISK1_4				0,744		
RISK1_5				0,778		
WILL1_1					0,929	
WILL1_2					0,961	
WILL1_3					0,949	
SENS_1						1

F.2 Factor loadings after modification

	PTRST	UTRST1	CTRL1	RISK1	WILL1	SENS
PTRST_1	0,722					
PTRST_2	0,86					
PTRST_3	0,897					
PTRST_4	0,776					
UTRST1_1		0,914				
UTRST1_2		0,952				
UTRST1_3		0,948				
CTRL1_1			0,809			
CTRL1_2			-			
CTRL1_3			0,917			
CTRL1_4			0,905			
RISK1_1				0,822		
RISK1_2				-		
RISK1_3				0,734		
RISK1_4				0,755		
RISK1_5				0,787		
WILL1_1					0,929	
WILL1_2					0,961	
WILL1_3					0,949	
SENS_1						1

F.3 Model fit

	Chi-square	Chi-Square/DF	CFI	TLI	RMSEA
Model 1	380,769	2,998181	0,943	0,947	0,082
Model 1 robust	308,281	2,427409	0,931	0,936	0,07

F.4 Regression paths

	β	Std.Err	Z-Value	P
PTRST ->CTRL1	0,244	0,145	3,444	0,001
CTRL1 ->RISK1	-0,367	0,053	-5,294	0
UTRST1 ->RISK1	-0,477	0,047	-7,801	0
SENS ->RISK1	0,258	0,041	4,966	0
PTRST ->RISK1	0,031	0,107	0,46	0,645
RISK1 - >WILL1	-0,673	0,09	-9,885	0

F.5 Modifications to PTRST

		Model 1		PTRST ->CTRL removed		PTRST ->CTRL reversed		PTRST ->RISK removed		PTRST removed	
	Chi square/DF	2,427		1,820		2,108		3,638		2,057	
			P		**P**		**P**		**P**		**P**
H1	**CTRL1 ->RISK1**	-0,405	0,000	-0,342	0,000	-0,344	0,000	-0,355	0,000	-0,343	0,000
H2	**UTRST1 ->RISK1**	-0,470	0,000	-0,434	0,000	-0,428	0,000	-0,406	0,000	-0,425	0,000
H3	**PTRST ->CTRL1**	0,253	0,000					0,192	0,005		
H3*	**CTRL1 ->PTRST**					0,240	0,000				
H4	**PTRST ->RISK1**	0,014	0,827	0,020	0,734	0,018	0,758				
H5	**SENS ->RISK1**	0,263	0,000	0,245	0,000	0,244	0,000	0,243	0,000	0,245	0,000
H6	**RISK1 - >WILL1**	-0,704	0,000	-0,694	0,000	-0,695	0,000	-0,676	0,000	-0,695	0,000

G Model 2 results

G.1 Factor loadings

	PTRST	**UTRST2**	**CTRL2**	**RISK2**	**WILL2**	**SENS**
PTRST_1	0,728					
PTRST_2	0,864					
PTRST_3	0,888					
PTRST_4	0,784					
UTRST2_1		0,866				
UTRST2_2		0,932				
UTRST2_3		0,952				
CTRL2_1			0,745			
CTRL2_2			0,775			
CTRL2_3			0,799			
CTRL2_4			0,75			
RISK2_1				0,807		
RISK2_2				0,783		
RISK2_3				0,865		
RISK2_4				0,707		
RISK2_5				0,749		
WILL2_1					0,932	
WILL2_2					0,966	
WILL2_3					0,912	
SENS_1						1

G.2 Factor loadings after modification

	PTRST	**UTRST2**	**CTRL2**	**RISK2**	**WILL2**	**SENS**
PTRST_1	0,728					
PTRST_2	0,864					
PTRST_3	0,888					
PTRST_4	0,784					
UTRST2_1		0,865				
UTRST2_2		0,932				
UTRST2_3		0,952				
CTRL2_1			0,609			
CTRL2_2			-			
CTRL2_3			0,898			
CTRL2_4			0,821			
RISK2_1				0,78		
RISK2_2				-		
RISK2_3				0,813		
RISK2_4				0,749		
RISK2_5				0,974		
WILL2_1					0,933	
WILL2_2					0,964	
WILL2_3					0,914	
SENS_1						1

G.3 Model fit

	Chi-square	Chi-Square/DF	CFI	TLI	RMSEA
Model 2	446,728	3,517543	0,922	0,906	0,093
Model 2 robust	301,613	2,374906	0,937	0,924	0,069

G.4 Regression paths

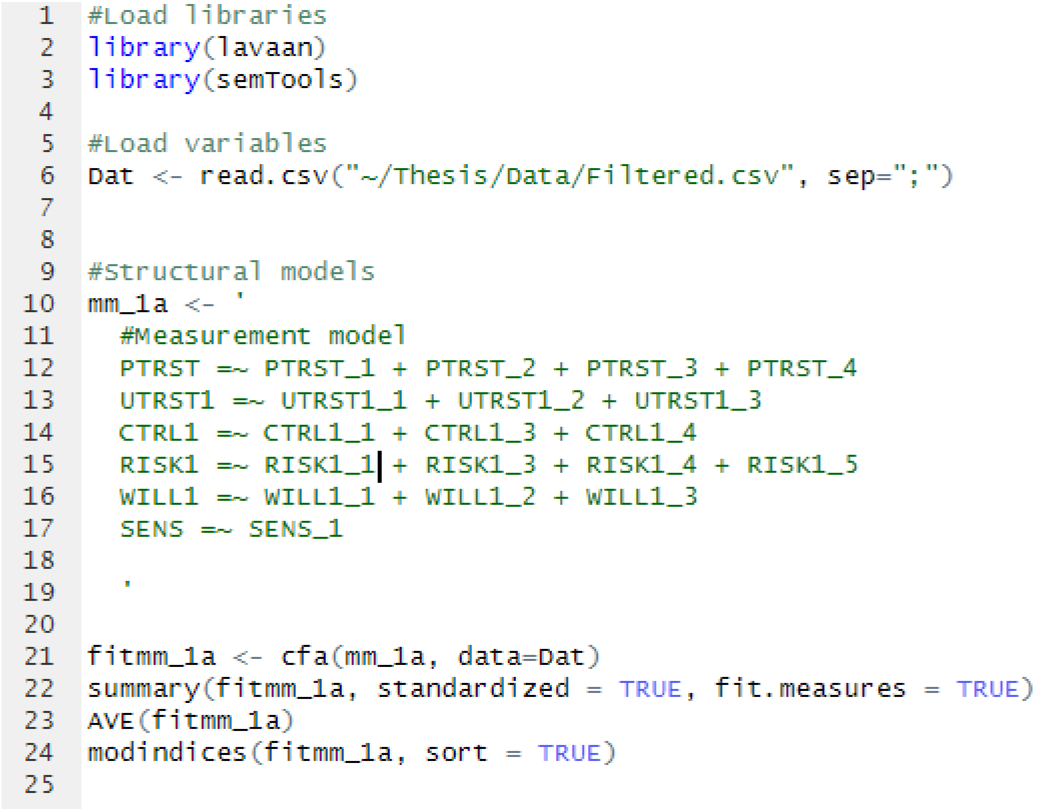
	β	Std.Err	Z-Value	P
PTRST ->CTRL2	0,474	0,102	5,212	0,000
CTRL2 ->RISK2	-0,515	0,106	-6,372	0,000
UTRST2 ->RISK2	-0,362	0,053	-6,259	0,000
SENS ->RISK2	0,196	0,038	3,855	0,000
PTRST ->RISK2	-0,016	0,105	-0,222	0,824
RISK2 - >WILL2	-0,747	0,085	-10,941	0,000

G.5 Modifications to PTRST

		Model 2		PTRST ->CTRL removed		PTRST ->CTRL reversed		PTRST ->RISK removed		PTRST removed	
	Chi square/DF	2,375		1,766		1,890		2,363		2,486	
			P		**P**		**P**		**P**		**P**
H1	**CTRL2 ->RISK2**	-0,515	0,000	-0,516	0,000	-0,559	0,000	-0,519	0,000	-0,519	0,000
H2	**UTRST2 ->RISK2**	-0,362	0,000	-0,264	0,001	-0,233	0,006	-0,370	0,000	-0,286	0,000
H3	**PTRST ->CTRL2**	0,474	0,000					0,467	0,000		
H3*	**CTRL2 ->PTRST**					0,485	0,000				
H4	**PTRST ->RISK2**	-0,016	0,824	-0,044	0,388	-0,019	0,738				
H5	**SENS ->RISK2**	0,196	0,000	0,182	0,000	0,180	0,000	0,197	0,000	0,183	0,000
H6	**RISK2 - >WILL2**	-0,747	0,000	-0,769	0,000	-0,769	0,000	-0,746	0,000	-0,767	0,000

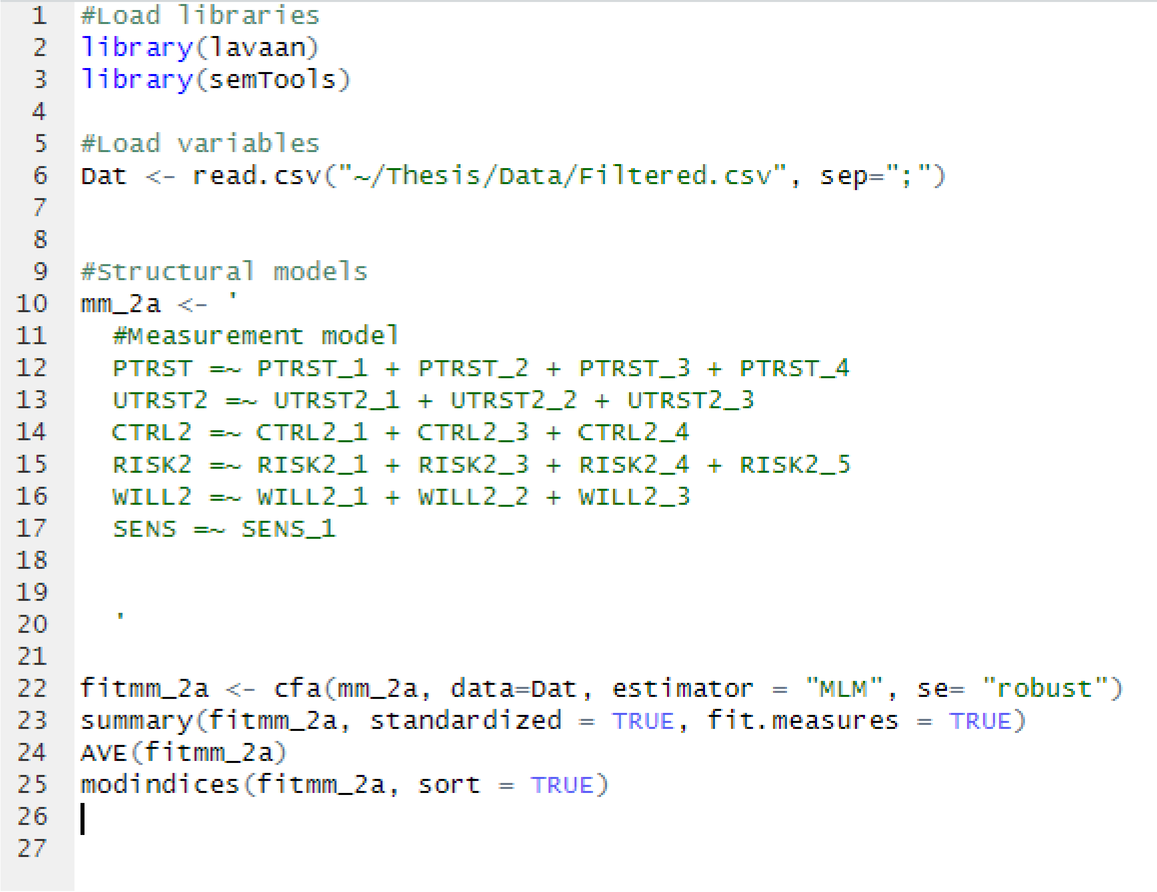
H R scripts

H.1 Measurement model 1



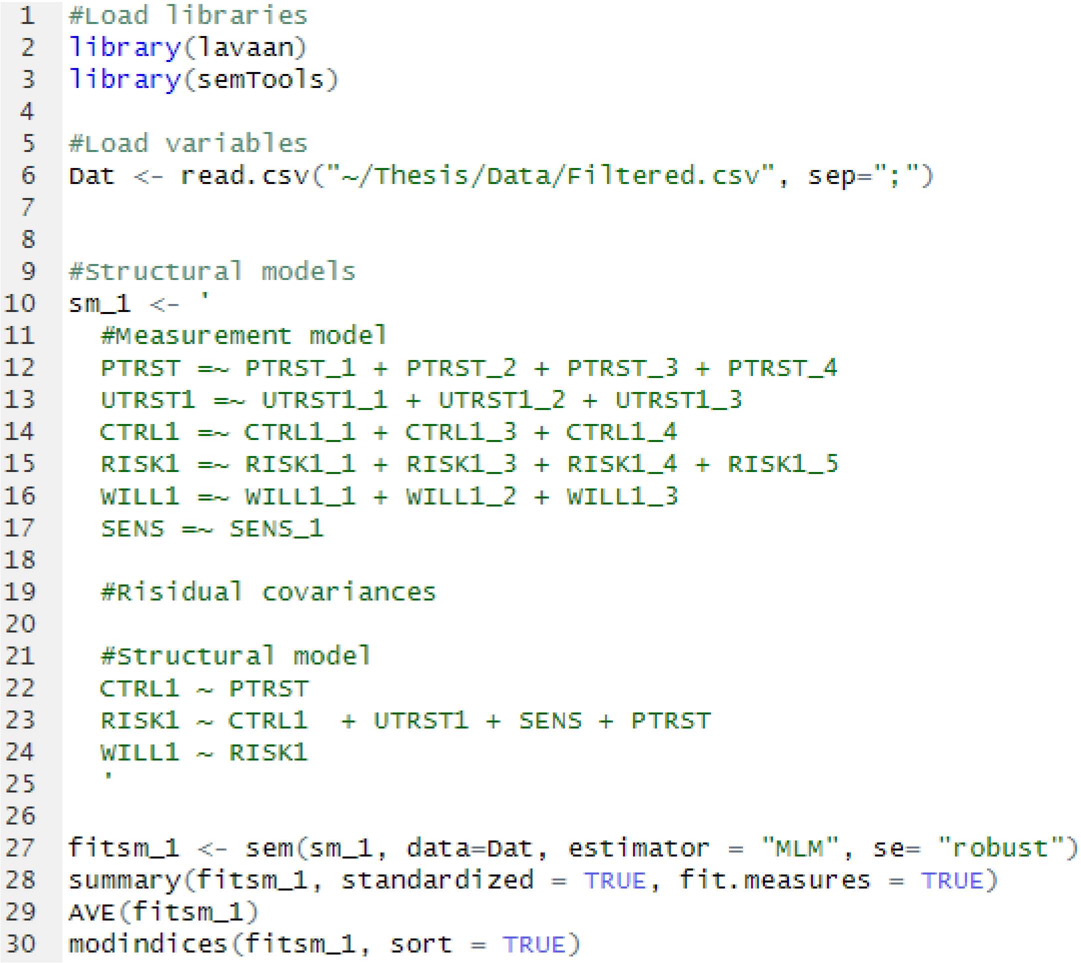
```r
#Load libraries
library(lavaan)
library(semTools)

#Load variables
Dat <- read.csv("~/Thesis/Data/Filtered.csv", sep=";")


#Structural models
mm_1a <- '
  #Measurement model
  PTRST =~ PTRST_1 + PTRST_2 + PTRST_3 + PTRST_4
  UTRST1 =~ UTRST1_1 + UTRST1_2 + UTRST1_3
  CTRL1 =~ CTRL1_1 + CTRL1_3 + CTRL1_4
  RISK1 =~ RISK1_1 + RISK1_3 + RISK1_4 + RISK1_5
  WILL1 =~ WILL1_1 + WILL1_2 + WILL1_3
  SENS =~ SENS_1

  '

fitmm_1a <- cfa(mm_1a, data=Dat)
summary(fitmm_1a, standardized = TRUE, fit.measures = TRUE)
AVE(fitmm_1a)
modindices(fitmm_1a, sort = TRUE)

```

H.2 Measurement model 2



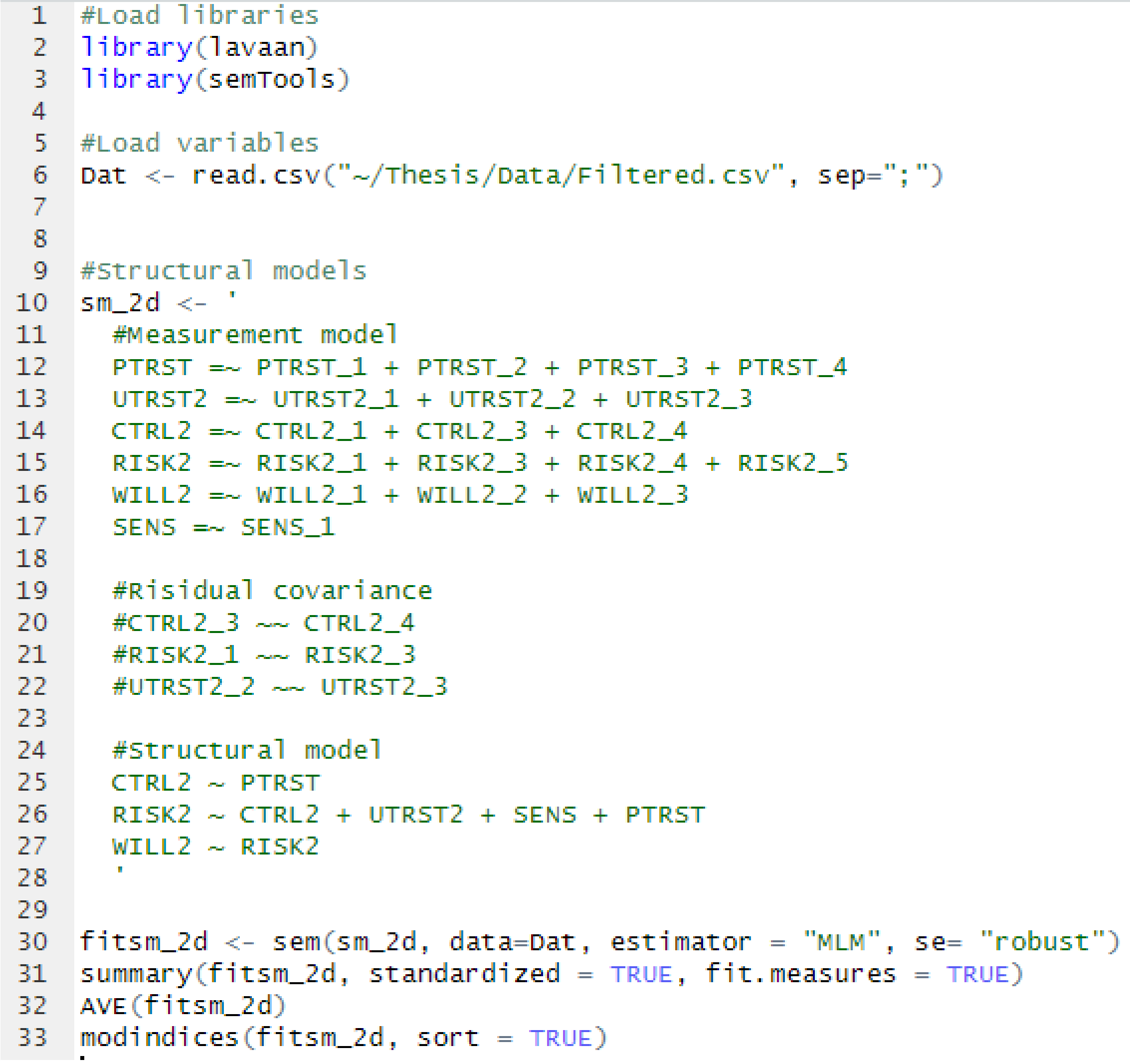
```
#Load libraries
library(lavaan)
library(semTools)

#Load variables
Dat <- read.csv("~/Thesis/Data/Filtered.csv", sep=";")


#Structural models
mm_2a <- '
  #Measurement model
  PTRST =~ PTRST_1 + PTRST_2 + PTRST_3 + PTRST_4
  UTRST2 =~ UTRST2_1 + UTRST2_2 + UTRST2_3
  CTRL2 =~ CTRL2_1 + CTRL2_3 + CTRL2_4
  RISK2 =~ RISK2_1 + RISK2_3 + RISK2_4 + RISK2_5
  WILL2 =~ WILL2_1 + WILL2_2 + WILL2_3
  SENS =~ SENS_1


  '

fitmm_2a <- cfa(mm_2a, data=Dat, estimator = "MLM", se= "robust")
summary(fitmm_2a, standardized = TRUE, fit.measures = TRUE)
AVE(fitmm_2a)
modindices(fitmm_2a, sort = TRUE)

```

H.3 Model 1



```r
#Load libraries
library(lavaan)
library(semTools)

#Load variables
Dat <- read.csv("~/Thesis/Data/Filtered.csv", sep=";")


#Structural models
sm_1 <- '
  #Measurement model
  PTRST =~ PTRST_1 + PTRST_2 + PTRST_3 + PTRST_4
  UTRST1 =~ UTRST1_1 + UTRST1_2 + UTRST1_3
  CTRL1 =~ CTRL1_1 + CTRL1_3 + CTRL1_4
  RISK1 =~ RISK1_1 + RISK1_3 + RISK1_4 + RISK1_5
  WILL1 =~ WILL1_1 + WILL1_2 + WILL1_3
  SENS =~ SENS_1

  #Risidual covariances

  #Structural model
  CTRL1 ~ PTRST
  RISK1 ~ CTRL1  + UTRST1 + SENS + PTRST
  WILL1 ~ RISK1
  '

fitsm_1 <- sem(sm_1, data=Dat, estimator = "MLM", se= "robust")
summary(fitsm_1, standardized = TRUE, fit.measures = TRUE)
AVE(fitsm_1)
modindices(fitsm_1, sort = TRUE)
```

H.4 Model 2



```
#Load libraries
library(lavaan)
library(semTools)

#Load variables
Dat <- read.csv("~/Thesis/Data/Filtered.csv", sep=";")


#Structural models
sm_2d <- '
  #Measurement model
  PTRST =~ PTRST_1 + PTRST_2 + PTRST_3 + PTRST_4
  UTRST2 =~ UTRST2_1 + UTRST2_2 + UTRST2_3
  CTRL2 =~ CTRL2_1 + CTRL2_3 + CTRL2_4
  RISK2 =~ RISK2_1 + RISK2_3 + RISK2_4 + RISK2_5
  WILL2 =~ WILL2_1 + WILL2_2 + WILL2_3
  SENS =~ SENS_1

  #Risidual covariance
  #CTRL2_3 ~~ CTRL2_4
  #RISK2_1 ~~ RISK2_3
  #UTRST2_2 ~~ UTRST2_3

  #Structural model
  CTRL2 ~ PTRST
  RISK2 ~ CTRL2 + UTRST2 + SENS + PTRST
  WILL2 ~ RISK2
  '

fitsm_2d <- sem(sm_2d, data=Dat, estimator = "MLM", se= "robust")
summary(fitsm_2d, standardized = TRUE, fit.measures = TRUE)
AVE(fitsm_2d)
modindices(fitsm_2d, sort = TRUE)
```

I Grouped analysis

Original models

	N	Model fit	Model fit/df	CFI	TLI	RMSEA
Model 1	295	308	2,427	0,931	0,936	0,070
Model 2	295	302	2,375	0,937	0,924	0,069
	Beta	**P**			**Beta**	**P**
PTRST -> CTRL1	0,253	0		PTRST -> CTRL2	0,474	0
CTRL1 -> RISK1	-0,405	0		CTRL2 -> RISK2	-0,515	0
UTRST1 -> RISK1	-0,47	0		UTRST2 -> RISK2	-0,362	0
SENS -> RISK1	0,263	0		SENS -> RISK2	0,196	0
PTRST -> RISK1	0,014	0,827		PTRST -> RISK2	-0,016	0,824
RISK1 -> WILL1	-0,704	0		RISK2 - > WILL2	-0,747	0

Familiarity data marketplaces > 3

	N	Model fit	Model fit/df	CFI	TLI	RMSEA	
Model 1	78	166	1,307	0,959	0,951	0,072	
Model 2	78	167	1,315	0,940	0,928	0,076	
	Beta	**P**	**Increase**		**Beta**	**P**	**Increase**
PTRST -> CTRL1	0,369	0,004	46%	PTRST -> CTRL2	0,414	0,033	-13%
CTRL1 -> RISK1	-0,516	0	27%	CTRL2 -> RISK2	-0,494	0	-4%
UTRST1 -> RISK1	-0,484	0	3%	UTRST2 -> RISK2	-0,333	0	-8%
SENS -> RISK1	0,288	0,002	10%	SENS -> RISK2	0,035	0,73	
PTRST -> RISK1	0,042	0,721		PTRST -> RISK2	0,152	0,224	
RISK1 - > WILL1	-0,684	0	-3%	RISK2 - > WILL2	-0,567	0	-24%

Familiarity IoT < 3

	N	Model fit	Model fit/df	CFI	TLI	RMSEA	
Model 1	109	200	1,571	0,946	0,935	0,078	
Model 2	109	210	1,654	0,898	0,877	0,091	
	Beta	**P**	**Increase**		**Beta**	**P**	**Increase**
PTRST ->CTRL1	0,0096	0,3		PTRST -> CTRL2	0,495	0	4%
CTRL1 ->RISK1	-0,483	0	19%	CTRL2 -> RISK2	-0,453	0	-12%
UTRST1 ->RISK1	-0,428	0	-9%	UTRST2 -> RISK2	-0,429	0	19%
SENS ->RISK1	0,36	0	37%	SENS -> RISK2	0,23	0,001	17%
PTRST ->RISK1	-0,166	0,04		PTRST -> RISK2	-0,094	0,342	
RISK1 - >WILL1	-0,768	0	9%	RISK2 - >WILL2	-0,832	0	11%

Sensitivity < 4

	N	Model fit	Model fit/df	CFI	TLI	RMSEA	
Model 1	62	160	1,262	0,947	0,937	0,065	
Model 2	62	188	1,483	0,872	0,846	0,106	
	Beta	**P**	**Increase**		**Beta**	**P**	**Increase**
PTRST ->CTRL1	0,179	0,146		PTRST -> CTRL2	0,55	0,099	16%
CTRL1 ->RISK1	-0,167	0,256		CTRL2 -> RISK2	-0,878	0,003	70%
UTRST1 ->RISK1	-0,543	0	16%	UTRST2 -> RISK2	0,042	0,825	
SENS ->RISK1	0,297	0,012	13%	SENS -> RISK2	0,214	0,082	9%
PTRST ->RISK1	0,15	0,315		PTRST -> RISK2	0,061	0,796	
RISK1 - >WILL1	-0,704	0	0%	RISK2 - >WILL2	-0,75	0	0%

Sensitivity > 4

	N	**Model fit**	**Model fit/df**	**CFI**	**TLI**	**RMSEA**	
Model 1	183	269	2,121	0,938	0,925	0,078	
Model 2	183	264	2,075	0,923	0,907	0,092	
	Beta	**P**	**Increase**		**Beta**	**P**	**Increase**
PTRST -> CTRL1	0,263	0,003	4%	PTRST -> CTRL2	0,426	0	-10%
CTRL1 -> RISK1	-0,447	0	10%	CTRL2 -> RISK2	-0,44	0	-15%
UTRST1 -> RISK1	-0,434	0	-8%	UTRST2 -> RISK2	-0,441	0	22%
SENS -> RISK1	0,249	0	-5%	SENS -> RISK2	0,103	0,114	
PTRST -> RISK1	-0,034	0,681		PTRST -> RISK2	-0,082	0,297	
RISK1 - > WILL1	-0,705	0	0%	RISK2 - > WILL2	-0,742	0	-1%

Middle managers

	N	**Model fit**	**Model fit/df**	**CFI**	**TLI**	**RMSEA**	
Model 1	142	236	1,854	0,943	0,931	0,850	
Model 2	142	204	1,606	0,954	0,944	0,076	
	Beta	**P**	**Increase**		**Beta**	**P**	**Increase**
PTRST -> CTRL1	0,25	0,014	-1%	PTRST -> CTRL2	0,49	0	3%
CTRL1 -> RISK1	-0,308	0,001	-24%	CTRL2 -> RISK2	-0,382	0	-26%
UTRST1 -> RISK1	-0,576	0	23%	UTRST2 -> RISK2	-0,456	0	26%
SENS -> RISK1	0,277	0	5%	SENS -> RISK2	0,128	0,041	-35%
PTRST -> RISK1	0,097	0,333		PTRST -> RISK2	-0,039	0,715	
RISK1 - > WILL1	-0,669	0	-5%	RISK2 - > WILL2	-0,751	0	1%

Years of experience > 5

	N	**Model fit**	**Model fit/df**	**CFI**	**TLI**	**RMSEA**	
Model 1	137	220	1,736	0,947	0,936	0,083	
Model 2	137	215	1,696	0,941	0,929	0,087	
	Beta	**P**	**Increase**		**Beta**	**P**	**Increase**
PTRST -> CTRL1	0,313	0,001	24%	PTRST -> CTRL2	0,654	0	38%
CTRL1 -> RISK1	-0,422	0	4%	CTRL2 -> RISK2	-0,464	0	-10%
UTRST1 -> RISK1	-0,485	0	3%	UTRST2 -> RISK2	-0,316	0	-13%
SENS -> RISK1	0,35	0	33%	SENS -> RISK2	0,235	0	20%
PTRST -> RISK1	0,15	0,15		PTRST -> RISK2	-0,086	0,555	
RISK1 - > WILL1	-0,693	0	-2%	RISK2 - > WILL2	-0,774	0	4%