Grant Agreement Number: 825225

Safe-DEED

www.safe-deed.eu

D2.7 User experiment report v3

Deliverable number	D2.7
Dissemination level	**Public**
Delivery date	30th November 2021
Status	Final
Author(s)	Wirawan Agahari, Mark de Reuver, Iris van de Wel, Christian van Aalst

This project has received funding from the European Union's Horizon 2020 research and innovation program (grant agreement no 825225).

Changes Summary

Date	Author	Summary	Version
7 September 2021	Mark de Reuver	Outline	0.1
8 October 2021	Wirawan Agahari	Draft of section 3 (study 2)	0.2
24 October 2021	Iris van der Wel	Draft of section 2 (study 1)	0.3
26 October 2021	Christian van Aalst	Draft of section 4 (study 3)	0.4
28 October 2021	Wirawan Agahari	First complete draft	1.0
2 November 2021	Mark de Reuver	Internal WP2 review	1.1
6 November 2021	Iris van der Wel	Layout, referencing, and study 1 revision	1.2
9 November 2021	Wirawan Agahari	Draft ready for review	2.0
16 November 2021	Panos Georgatsos, Evangelos Kotsifakos (LSTECH)	Final version review	2.1
16 November 2021	Leonie Disch	Final version review	2.2
17 November 2021	Wirawan Agahari	Process final version review	2.3
24 November 2021	Iris van der Wel	Process final version review	2.4
25 November 2021	Wirawan Agahari	Final version	2.5

Executive summary

The Safe-DEED project strives to enhance trust in the data economy by enabling the large-scale implementation of privacy-preserving technologies. Ultimately, the goal is to overcome data sharing barriers and ultimately accelerate the European data economy. The objective of task T2.3 is to measure the impact of Safe-DEED technologies on citizen trust and Willingness to Share data as a key to unlocking the data economy potential.

The first deliverable (D2.4) of T2.3 evaluated a mock-up based on WP7, and the second deliverable (D2.6) focused on Multi-Party Computation (MPC) as developed in WP5. In this third and last deliverable of T2.3, our main objective is to evaluate the three final version demonstrators as developed in WP4, WP5, and WP6. We consider criteria of trust, complexity, security, control, benefit, and intention to use the demonstrators. Through a large-scale online survey, we study the expectations and use experiences. We find that the actual perception of demonstrators was largely positive on all criteria considered. Surprisingly, trust and security are not significant factors that influence the willingness to use. This is probably due to the context of the demonstrator that is not very threatening, making trust and security less important than benefits, complexity, and control. People who have more experience and knowledge in data analytics have a more positive impression about the demonstrators than inexperienced people – likely because they are able to see and appreciate the value of the demonstrator better.

In two follow-up studies, we evaluate underlying privacy-preserving technologies in the Safe-DEED demonstrator within the context of personal data marketplaces. We first compare the decentralized data sharing through data marketplaces based on MPC to the current state-of-the-art way of centralized data sharing with a trusted third party. We focus on values of control, Perceived Risk, privacy concerns, trust, and intention to share data. Then, we assess the relative importance of where MPC algorithms are deployed (centralized or decentralized) compared to aspects like the risk of data leakage, social influence, and monetary benefits of selling data. Results indicate that Safe-DEED technologies would make people more willing to share data than sharing through trusted third parties, as people perceive a stronger feeling of control and trust as well as lower risk perception and privacy concerns because of MPC. Nevertheless, people do not really care about how MPC is deployed and put more weight on risks, benefits, and social influence.

We infer from our findings that Safe-DEED technologies could contribute toward trust and intention to share data of both businesses and individuals. The demonstrators receive positive evaluations on all of the criteria set out to achieve in the project. The Safe-DEED approach to decentralized data collaboration is also found to outperform traditional data sharing through trusted third parties. The technologies, particularly MPC, could even create more value beyond trust by inducing the feeling of control over data, reducing the risk of data sharing, and lowering privacy concerns. This is possible since only the computation results will be shared with other parties, not the input data. Data owners could have more control over how their data is used, while data users would still be able to generate value from the computation results without harming the privacy of data owners.

To ensure that Safe-DEED technologies could deliver those values as intended, we recommend (1) developing ways to communicate and visualize how the technology works and its potential benefits for users; (2) emphasizing potential users that are knowledgeable and highly engaged in data-related activities as a target group; and (3) complement those with proper data governance mechanisms that go beyond technical solution.

Tables of Content

List of Figures

List of Tables

1. Introduction

Safe-DEED aims to advance the data economy through technologies that facilitate data sharing in privacy and confidentiality preserving ways. These technologies are being developed, prototyped, and demonstrated in WP4 and WP5, with the help of use cases in WP6 and WP7. In WP2, one of the objectives is to evaluate whether the developed technologies contribute to fulfilling the goals of the data economy. Specifically, we consider how the developed technologies contribute to citizen trust in the data economy, value creation for businesses, adoption of privacy/confidentiality preserving technologies, and ultimately turnover of businesses.

This deliverable is the final one in a series of deliverables that report user experiments in T2.3. In the first deliverable (D2.4), we evaluated a mock-up created within WP5 and WP7 and found that privacy-preserving technologies, specifically MPC, affect user trust, Perceived Security, and intention to share data depending on how MPC is communicated to the user. In the second deliverable (D2.6), we evaluated mock-ups of MPC-enabled data sharing in supply chains. In this third and final deliverable, we evaluate how the developed prototypes and demonstrators of WP4, WP5, and WP6 affect usefulness, ease-of-use, and intention to use. In contrast to the first two deliverables, we mainly focus on the citizen perspective of data sharing rather than business perspectives.

We do so through three studies. In the first and main study, we conduct a large-scale online survey in which participants are asked to try out the three demonstrators[1] (data exchange application, de-anonymization risk analysis, and data valuation technologies). We then assess user perceptions, including the intention to use the demonstrator and compare these to participants' expectations before trying out the demonstrator. The results provide a direct evaluation of the developed technologies in Safe-DEED.

While the first study provides insights into the extent to which users perceive the Safe-DEED demonstrators positively, it only focuses on a specific context of data sharing between businesses. This means that whether the developed technology could also outperform existing data sharing approaches in different settings, especially consumer-to-business data sharing, remains unclear. In other words, we do not know for sure how Safe-DEED technologies could affect citizens' trust and privacy-preservation in general. Filling this gap is the key focus of T2.3, and therefore, we conducted two follow-up studies that focus on consumers' perspectives in data sharing within the specific context of personal data marketplaces. Study 2 compares Safe-DEED's approach of decentralized data sharing to the current state-of-the-art way of centralized data sharing. We compare how these different approaches perform regarding perceived risks, trust, and intention to share data. In study 3, we assess and compare the relative importance of MPC deployment scenarios (i.e., centralized versus decentralized architecture) for citizens to aspects of risk of data leakage, social influence, and monetary benefits of selling data. See Table 1 for an overview of the three studies.

[1] For a detailed explanation about the demonstrator, please refer to D4.5 and D6.3.

Study	Context	Approach	Key questions
Study 1: Evaluation of Safe-DEED demonstrators	Business-to-Business data sharing	Pre-test-post-test within-subject experimental design	Would business actors be willing to use the Safe-DEED demonstrator?
Study 2: Comparison to trusted third party scenario	Consumer-to-Business data sharing in the automotive context	Post-test only between-subject experimental design	Does MPC give more feeling of trust, control and ultimately make consumers more willing to share their driving data?
Study 3: Identifying consumers' preference in sharing driving data on MPC-enabled data marketplaces	Consumer-to-Business data sharing in the automotive context	Discrete choice experiment	What are the consumers' preferences concerning MPC architecture in data marketplaces?

Table 1 Comparison between three studies

We collected data for all studies using the online crowdsourcing platform Prolific (Palan & Schitter, 2018). Such a platform allows us to collect a large number of survey responses in a short time and is commonly used in academic research nowadays. Beyond that, compared to Amazon Mechanical Turk (Mturk), participants recruited via Prolific are more diverse, naïve, and honest (Adams, Li, & Liu, 2020; Peer, Brandimarte, Samat, & Acquisti, 2017). Nevertheless, we are aware of disadvantages in using Prolific for academic research, such as selection bias by participants and monetary incentives (Kaufmann, Schulze, & Veit, 2011). In each study, we outlined a detailed process in selecting our samples.

The following sections explain in more detail how we conducted each study in Section 2 (Study 1), Section 3 (Study 2), and Section 4 (Study 3). We begin by explaining research approaches and the rationale behind those, followed by measures used in each study and the profile of our samples. After that, we present the results and specific conclusions of each study. Finally, we integrate our findings and discuss its implication for the Safe-DEED project in section 5.

2. Study 1: Evaluation of demonstrators

WP4, WP5, and WP6 developed demonstrators for different data-sharing applications that are based on privacy-preserving technologies, namely: a data exchange application, a de-anonymization application, and a data valuation application[2]. This study examines all three demonstrators to provide insights into how users perceive these technologies concerning their benefits, complexity, control, security, trustworthiness, and willingness to use. To fulfill this objective, we perform three experiments that measure the expectations for certain applications and that measure the actual perception of the demonstrators. This could give us meaningful information about how the applications developed in Safe-DEED are perceived by business users and to what extent this differs from what they expect. As a unit of analysis, we scope our research from the perspective of business actors.

Section 2.1 elaborates on the used method, the experiment setup, and the sample we used. After that, we discuss the results of the statistical test in section 2.2. Finally, we elaborate on insights derived from the survey data and conclude our study in section 2.3.

2.1 Method

2.1.1 Experimental design

We designed an online experiment to measure the influence of the data exchange, de-anonymization, and data valuation demonstrators on perceived benefit, perceived complexity, perceived control, perceived security, perceived trustworthiness, and willingness to use. We did so through an independent pre-test post-test experiment for the data exchange application, de-anonymization application, and data valuation application, which is often used to measure the influence of treatment on behavior. In our case, the demonstrators can be seen as the treatments (Dimitrov & Rumrill, 2003). This experimental design allowed us to measure the influence of the demonstrator on the willingness to use data sharing applications. Instead of measuring behavior, we measured the difference between expectations of an application and how the demonstrators were perceived. Precisely, we measured the expectations for an application in a specific scenario in the pre-test. Then, the demonstrator was introduced and explained in detail. The post-test measured the perception of the application with the constructs defined in section 2.1.2. The terms application and demonstrator are used interchangeably in this report.

Specifically, a one-group pre-test post-test design was used in this research, meaning that there is only a test group wherein the same participants get the pre- and post-test. This can also be defined as a within-subject design, as all participants get the same treatment and the conditions for them are the same in each experiment (Charness, Gneezy & Kuhn, 2012). We opt for this design because it is easy to implement and analyze (Cranmer, 2018) and thereby advantageous for our research that includes three experiments. It should be noted that this quasi-experimental design has been criticized for not having a control group, which could be problematic for the internal validity of the study. To minimize this problem, the dependent factors were kept constant for the pre-test and post-test (Privitera & Ahlgrim-Delzell, 2018).

[2] These demonstrators were explained in detail in D4.5 and D6.3

Experiment	Pre-test	Treatment	Post-test
1	O_1	$X_{\text{data exchange application}}$	O_2
2	O_1	$X_{\text{de-anonymization application}}$	O_2
3	O_1	$X_{\text{data valuation application}}$	O_2

Table 2 Experimental design for study 1

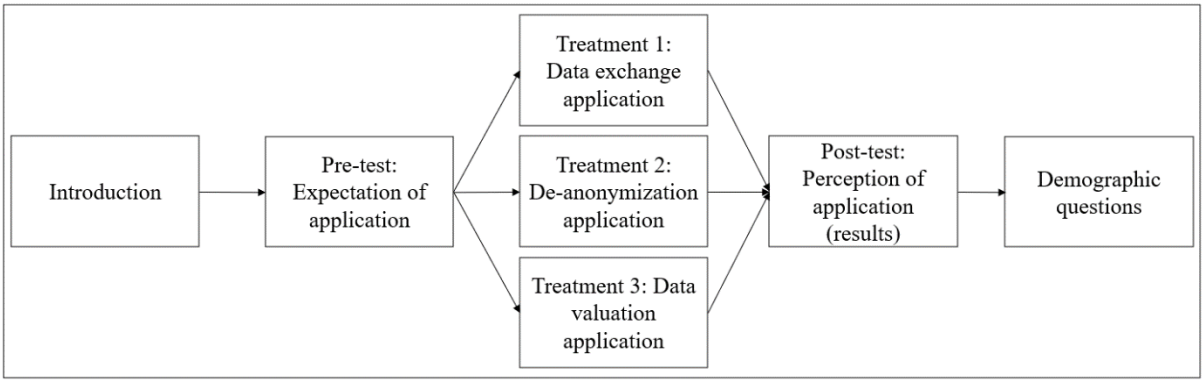


Figure 1 Survey flow

We performed experiments for three demonstrators based on privacy-preserving technologies. The experimental setup is shown in Table 2. The experiments are conducted through online surveys. Figure 1 presents the flow of these surveys. Experiment 1 examined the data exchange application that was developed by WP5. This application uses a Private Set Intersection (PSI) protocol that computes the intersection of two datasets from different owners without releasing the underlying data. Experiment 2 examined the de-anonymization application developed by WP5. This application can check the de-anonymization risks of a dataset and can anonymize data through a k-anonymization technique. Experiment 3 examined the data valuation application developed by WP4. This application helps with assessing the quality, exploitability, and economic value of data. The application works by receiving a structured data set, together with a context and a set of rules for evaluating data quality.

Next, we present the scenarios that were used to explain the demonstrators and underlying techniques to the participants. It is likely that most of our respondents have not heard of the privacy-preserving technologies developed by Safe-DEED. The use of scenarios contributes to a better understanding of the possibilities of a certain technology. In this scenario, we introduced participants with a persona in which they are a business development manager of a telecom operator selling premium pay-TV packages. Their company is a market leader that is successful in selling live sports TV packages. However, they would like to increase revenues and subscriber base and are therefore interested in exploring Safe-DEED technologies to achieve this goal.

The next part describes different situations depending on the experiment. In experiment 1, the manager would like to achieve the goal above by identifying cross-selling opportunities with a bank. The data exchange application is introduced as a solution for sharing confidential information. In experiment 2, the manager would like to increase revenues by analysing privacy-sensitive data. The de-anonymization application is introduced as a solution to check whether the dataset is sufficiently anonymized and to anonymize the data. In experiment 3, the manager would like to increase revenues by analysing usage patterns of live streams data. The manager wants to assess its value and quality. The data valuation application then offers the solution by calculating the qualitative data score, an automatic data analysis score, and a dataset value.

In each experiment, we asked participants to perform several tasks to engage with the demonstrator. In experiment 1, participants needed to upload datasets containing all the zipcodes of the premium pay-TV customers and compare the results with datasets of another company through data intersection. In

experiment 2, participants had to upload demographic data of all the premium pay-TV customers and perform a de-anonymization risk analysis. Then, the data was anonymized, and we asked participants to compare the anonymized data with the original file. Finally, in experiment 3, we showed a video of how the data valuation application works due to the complex steps that participants need to do. The complete procedures of the experiments are described in Appendix A.

2.1.2 Measures

As described earlier, the constructs we measured in this study are **Perceived Benefits**, **Perceived Complexity**, **Perceived Control**, **Perceived Security**, **Perceived Trustworthiness**, and **Willingness to Use**. The importance of these variables and their aspects became apparent from D2.6. Therefore, we developed a 5-scale Likert questionnaire by modifying measures developed in D2.6 and adjusted those measures depending on our demonstrators (Petronia, 2020; Safe-DEED, 2020). The scale ranges from 1 (Totally disagree) to 5 (Totally agree). The items in the experiment of the data exchange application and the de-anonymization application are similar. **Perceived Security** and **Perceived Control** are excluded from experiment 3 concerning the data valuation application because the privacy and security issues apply to a lesser extent to this application. Where applicable, we use multiple items to measure constructs, as this allows assessing construct validity and reliability (see 2.1.4). See Table 3 for an overview of the constructs.

Constructs	Pre-test		Post-test	
	Item	**Description**	**Item**	**Description**
Perceived Trustworthiness	TW_E1	I expect the claims made by the application to be clear and accurate.	TW_R1	Claims made by the application are clear and accurate.
	TW_E2	I expect the [name demonstrator] process of the application to be trustworthy.*	TW_R2	The [name demonstrator] process is trustworthy.*
Perceived Complexity	CX_E1	I expect the purpose of the application to be clear to me.	CX_R1	The purpose of the application is clear to me.
	CX_E2	I expect the application to provide a complete and detailed description of how it works.	CX_R2	The application provides a complete and detailed description of how it works.
	CX_E3	I expect the descriptions of the application to be complex.	CX_R3	The application descriptions are complex.
	CX_E4	I expect the application to be easy to understand and easy to use.	CX_R4	It is easy to understand how it works and easy to use.
Perceived Security	SC_E	I expect the application to make me feel secure in using it.*	SC_R	It makes me feel secure in using it.*
Perceived Control	CT_E	I expect the application to make me feel in control over my data.*	CT_R	It makes me feel in control over my data.*
Perceived Benefit	BN_E1	I expect to be able to apply the results to value my company's data.**	BN_R1	I can apply the results to value my company's data.**
	BN_E2	I expect my company to benefit from using the application.	BN_R2	My company would get benefit from using it.

Willingness to Use	WU_E1	I expect the application to make me feel less hesitant in exchanging sensitive business data to external companies.*/ I expect knowing the value and quality of my data to make me feel less hesitant in exchanging sensitive business data to other parties.**	WU_R1	Using the [name demonstrator] makes me feel less hesitant in exchanging sensitive business data to other parties.
			WU_R2	I would be willing to use the [name demonstrator] within the next 6 months or 1 year.

** These items were not included in experiment 3 concerning the data valuation application.*

*** These items were solely included in experiment 3 concerning the data valuation application.*

Table 3 Measures of experiments 1, 2 and 3

2.1.3 Sampling

The population we studied are highly educated people with a (former) function at work in which they are able to make decisions about adopting new applications. These are likely the people who decide upon the use of the demonstrators in reality. The extent to which these individuals are involved in evaluation, adoption, or implementation are relevant factors to consider in our study (Petronia, 2020). We used two criteria to preselect participants in Prolific: occupation level and education level. The education level ('Highest education level completed') is set to Undergraduate degree (BA/BSc/other), Graduate degree (MA/MSc/Mphil/other), or Doctorate degree (Ph.D./other). The occupation level ('Industry Role') is set to Upper Management, Trained Professional, Middle Management, or Junior Management.

We launched the survey between 21 June 2021 – 21 July 2021. In total, 600 participants took part in the study, comprised of 200 participants for each experiment. The large sample size allows conducting of parametric tests. Table 4 shows the demographics of the participants of the three experiments. It can be concluded that the participants from the three experiments are similar. The three samples have mostly full-time working people, have an (under)graduate degree, and have a medium occupation level at work. Despite our selection criteria, there were people with a diploma from high school, community college, or secondary education in our sample. Hence, we removed these datasets and continued with the participants who met the required education level. Nevertheless, while we also have students in our sample (N=6 in experiment 1, N=4 in experiment 2, N=1 in experiment 3), we keep these datasets because they could have (part-time) jobs or companies in which they are able to make decisions and they do have at least an undergraduate degree.

It is hard to determine whether the required occupational level is met. We cannot conclude exactly whether people from an entry-level do or do not meet our criteria through our survey question regarding the participants' seniority level at work. An intern from an entry-level does not meet our criteria, but some roles in this level can be seen as junior management. We can conclude that the sample includes people from three different seniority levels, namely Entry-level (36.7%), Mid-level (48.5%), and Senior-level (10.2%). Also, we assume that the prolific users set up their profile characteristics in good faith and therefore meet our selection criteria.

Lastly, it should be noted that there are many respondents from Portugal in all experiments (27.6% in experiment 1, 23.8% in experiment 2, and 24.1% in experiment 3). One explanation could be the relatively high unemployment rate in Portugal in comparison to other countries from the European Union. In 2017 Portugal was the country with the seventh-highest unemployment rate (9%) (CBS, 2019). They may therefore have a greater need to earn money by completing online surveys. Another explanation for the background of the participants can be found in the time at which we published the surveys. This is likely to affect who takes the surveys due to a difference in time zones. When publishing

a survey in the middle of the day, as we did, people from the same time zone are more likely to fill it in. However, this does not explain why there are relatively many people from Mexico.

Demographic	Values	1. Data exchange app (N=196)		2. De-anonymization app (N=189*)		3. Data valuation app (N=187)	
		Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Gender	Male	121	61.7	128	67.7	109	58.3
	Female	75	38.3	61	32.3	78	41.7
Age (mean, Sd)		30.5 (7.2)		32.0 (8.0)		31.8 (7.9)	
Educational degree	Undergraduate (BA, BSc, other)	68	34.7	78	41.3	69	36.9
	Graduate degree (MA, MSc, other)	118	60.2	99	52.4	107	57.2
	Doctorate degree (PhD, other)	10	5.1	12	6.3	11	5.9
Country	Portugal	54	27.6	45	23.8	45	24.1
	Mexico	13	6.6	14	7.4	11	5.9
	Poland	20	10.2	17	9	16	8.6
	United Kingdom	18	9.2	31	16.4	23	12.3
	South Africa	16	8.2	13	6.9	28	15.0
	Others	75	38.3	69	36.5	64	34.2
Work	Working full-time	159	81.1	156	82.5	157	84.0
	Working part-time	29	14.8	28	14.8	28	15.0
	Student	6	3.1	4	2.1	1	0.5
	On leave but still employed	1	0.5	0	0	0	0.0
	Wanting to work, but unemployed due to personal reasons	1	0.5	0	0	0	0.0
	Missing values			1	0.5	1	0.5
Seniority level	Entry-level (e.g., intern, trainee, staff, associate)	72	36.7	58	30.7	59	31.6
	Mid-level (e.g., manager, supervisor, team leader)	95	48.5	110	58.2	102	54.5
	Senior-level (e.g., department head, vice president, director)	20	10.2	15	7.9	19	10.2
	Not applicable	9	4.6	6	3.2	2	1.1
	Others	0	0.0	0	0.0	5	2.7
Sector	Computing or IT	40	20.4	33	17.5	31	16.6
	Engineering or manufacturing	21	10.7	27	14.3	18	9.6

	Science or pharma-ceuticals	18	9.2	16	8.5	17	9.1
	Business, consultancy or management	17	8.7	12	6.3	14	7.5
	Others	100	51.0	101	53.4	107	57.2
Organization Size	1	1	0.5	1	0.5	4	2.1
	2-10	28	14.3	34	18	25	13.4
	11-100	60	30.6	55	29.1	60	32.1
	>100	105	53.6	96	50.8	94	50.3
	Missing values	2	1.0	3	1.6	4	2.1
Involvement in developing new products	Never	16	8.2	17	9	18	9.6
	Rarely	44	22.4	42	22.2	35	18.7
	Sometimes	69	35.2	69	36.5	69	36.9
	Often	48	24.5	50	26.5	46	24.6
	Always	19	9.7	11	5.8	19	10.2
Data-related role at work**	Not involved	58	29.6	52	27.5	60	32.1
	Slightly involved	61	31.1	52	27.5	49	26.2
	Somewhat involved	37	18.9	39	20.6	43	23
	Moderately involved	25	12.8	27	14.3	23	12.3
	Very involved	15	7.7	16	8.5	12	6.4
Familiarity with MPC/de-anony-mization/data valuation***	Not familiar	107	54.6	82	43.4	33	17.6
	Slightly familiar	49	25	44	23.3	51	27.3
	Somewhat familiar	20	10.2	36	19	54	28.9
	Moderately familiar	17	8.7	21	11.1	39	20.9
	Very familiar	3	1.5	3	1.6	10	5.3

** N=186 for a data-related role at work and familiarity with de-anonymization due to missing values.*

*** The question that was asked: to what extent are/were you involved in a data-related role at your work? Examples are data managers, data scientists, or data engineers.*

**** How familiar are you with [name demonstrator]?*

Table 4 Demographic characteristics

2.2 Results

2.2.1 Forming constructs

Before we analyzed the collected data, we formed constructs; these are the factors that consist of multiple items. Table 5 shows the factors of our experiments and the corresponding abbreviations. **Perceived Trustworthiness** and **Perceived Complexity** consist of multiple items. Before we summed the items, we checked the one-dimensionality, correlations, and reliability of the scales.

Construct	Abbreviation	Experiment
Perceived Trustworthiness	TW	1,2,3
Perceived Complexity	CX	1,2,3
Perceived Security	SC	1,2
Perceived Control	CT	1,2
Perceived Benefit	BN	1,2,3
Willingness to Use	WU	1,2,3

Table 5 Constructs from experiment 1-3

We performed a Principal Axis Factoring (PAF) analysis to determine the underlying latent factors and see if these correspond with the constructs we use. The communalities of all items are greater than 0.25 and should thus be included in the PAF analysis.

Table 6 shows the factor loadings of the items for the pre-test and post-test of the data exchange application after PAF. In the pre-test, a simple structure was achieved with varimax rotation. In the post-test, a simple structure was achieved without rotation. It can be noted that almost all items from different constructs load high on one factor. From the PAF analysis, it followed that there are two factors with an eigenvalue greater than one. One factor includes items from **Perceived Complexity** and **Perceived Trustworthiness**, and one factor includes item 3 from **Perceived Complexity**. This may mean this item did not measure the same as the other items from **Perceived Complexity**. Thus, there is some overlap between the scales of the constructs. It can be questioned whether the scales from D2.6 are applicable to measuring trust. Since an informed choice has been made to use these scales, this overlap is not expected to cause problems for further analyses.

1. Data exchange application					
Pre-test*			**Post-test****		
Item	1	2	Item	1	2
CX_E1	**0,667**	0,202	CX_R1	**0,580**	-0,085
CX_E2	**0,634**	-0,063	CX_R2	**0,713**	-0,058
CX_E3	-0,010	-0,392	CX_R3	-0,075	0,386
CX_E4	0,219	**0,811**	CX_R4	**0,685**	-0,148
TW_E1	**0,593**	0,196	TW_R1	**0,744**	0,038
TW_E2	0,409	0,372	TW_R2	**0,588**	0,329

* *Varimax rotation*

** *No rotation*

Table 6 Factor loadings in experiment 1

Table 7 shows the factor loadings of the items for the pre-test and post-test of the de-anonymization application after PAF. In both the pre-test and post-test, a simple structure was achieved without rotation. It is also noticeable here that almost all items load high on one factor instead of the two underlying latent factors.

1. De-anonymization application				
Pre-test*			**Post-test***	
Item	1	2	Item	1
CX_E1	**0,617**	0,246	CX_R1	**0,629**
CX_E2	0,422	0,409	CX_R2	**0,720**
CX_E3	-0,373	0,240	CX_R3	-0,394
CX_E4	**0,790**	-0,335	CX_R4	**0,731**
TW_E1	**0,534**	0,124	TW_R1	**0,724**
TW_E2	0,487	-0,077	TW_R2	**0,544**

** No rotation*

Table 7 Factor loadings in experiment 2

Table 8 shows the factor loadings of the items for the pre-test and post-test of the de-anonymization application after PAF with no rotation. In both the pre-test and post-test, a simple structure was achieved without rotation. It is also noticeable here that almost all items load high on one factor instead of the two underlying latent factors. Item 3 of **Perceived Complexity** (CX_E3) seems to measure something different than the other items of **Perceived Complexity**.

3. Data valuation application					
Pre-test*			**Post-test***		
Item	1	2	Item	1	2
CX_E1	**0,692**	0,259	CX_R1	0,359	0,481
CX_E2	**0,588**	0,196	CX_R2	0,410	**0,525**
CX_E3	-0,318	**0,556**	CX_R3	-0,018	-0,305
CX_E4	**0,634**	-0,270	CX_R4	0,215	**0,650**
BN_E1	**0,638**	0,084	BN_R1	**0,922**	-0,354
BN_E2	**0,768**	0,000	BN_R2	**0,763**	-0,271

** No rotation*

Table 8 Factor loadings in experiment 3

Next, we performed correlation analyses. The correlations of all items of a construct should be positive; otherwise, it will cause problems in making the sum scores. Table 9-Table 14 shows the correlation of **Perceived Complexity**, **Perceived Trustworthiness**, and **Perceived Benefit**. Item 3 of **Perceived Complexity** (CX_E3) has negative correlations with the other items from **Perceived Benefit**. Besides, it is often low, which confirms the finding that this item might measure something else. CX_E3 states as follows: 'I expect the descriptions of the application to be complex.'. The statement was indeed formulated negatively. Therefore, it needed to be recoded. The reliability analysis will confirm whether the item needs to be deleted from the scale.

Data exchange application									
Pre-test					**Post-test**				
Item	CX_E1	CX_E2	CX_E3	CX_E4	**Item**	CX_R1	CX_R2	CX_R3	CX_R4
CX_E1	1.000				**CX_R1**	1.000			
CX_E2	0.433	1.000			**CX_R2**	0.401	1.000		
p	(*0.000*)				***p***	(*0.000*)			
CX_E3	-0.152	-0.054	1.000		**CX_R3**	-0.060	-0.077	1.000	
p	(*0.038*)	(0.467)			***p***	(0.400)	(0.285)		
CX_E4	0.295	0.102	-0.343	1.000	**CX_R4**	0.405	0.514	-0.119	1.000
p	(*0.038*)	(0.165)	(*0.000*)		***p***	(*0.000*)	(0.000)	(0.098)	

Table 9 Correlations of Perceived Complexity in experiment 1

Data exchange application					
Pre-test			Post-test		
Item	TW_E1	TW_E2	**Item**	TW_R1	TW_R2
TW_E1	1.000		**TW_R1**	1.000	
TW_E2	0.377	1.000	**TW_R2**	0.453	1.000
p	(*0.000*)		***p***	(*0.000*)	

Table 10 Correlations of Perceived Trustworthiness in experiment 1

De-anonymization application									
Pre-test					Post-test				
Item	CX_E1	CX_E2	CX_E3	CX_E4	**Item**	CX_R1	CX_R2	CX_R3	CX_R4
CX_E1	1.000				**CX_R1**	1.000			
CX_E2	0.365	1.000			**CX_R2**	0.404	1.000		
p	(*0.000*)				***p***	(*0.000*)			
CX_E3	-0.106	-0.081	1.000		**CX_R3**	-0.217	-0.282	1.000	
p	(0.151)	(0.274)			***p***	(*0.003*)	(*0.000*)		
CX_E4	0.457	0.177	-0.390	1.000	**CX_R4**	0.456	0.542	-0.452	1.000
p	(*0.000*)	(0.016)	(*0.000*)		***p***	(*0.000*)	(*0.000*)	(*0.000*)	

Table 11 Correlations of Perceived Complexity in experiment 2

De-anonymization application					
Pre-test			Post-test		
Item	TW_E1	TW_E2	**Item**	TW_R1	TW_R2
TW_E1	1.000		**TW_R1**	1.000	
TW_E2	0.310	1.000	**TW_R2**	0.455	1.000
p	(*0.000*)		***p***	(*0.000*)	

Table 12 Correlations of Perceived Trustworthiness in experiment 2

Data valuation application									
Pre-test					Post-test				
Item	CX_E1	CX_E2	CX_E3	CX_E4	**Item**	CX_R1	CX_R2	CX_R3	CX_R4
CX_E1	1.000				**CX_R1**	1.000			
CX_E2	0.493	1.000			**CX_R2**	0.536	1.000		
p	(*0.000*)				***p***	(*0.000*)			
CX_E3	-0.093	-0.040	1.000		**CX_R3**	-0.248	-0.184	1.000	
p	(0.151)	(0.593)			***p***	(*0.001*)	(*0.012*)		
CX_E4	0.299	0.411	-0.372	1.000	**CX_R4**	0.561	0.566	-0.377	1.000
p	(*0.000*)	(*0.000*)	(*0.000*)		***p***	(*0.000*)	(*0.000*)	(*0.000*)	

Table 13 Correlations of Perceived Complexity in experiment 3

Data valuation application					
Pre-test			**Post-test**		
Item	BN_E1	BN_E2	**Item**	BN_R1	BN_R2
BN_E1	1.000		**BN_R1**	1.000	
p			***p***		
BN_E2	0.555	1.000	**BN_R2**	0.805	1.000
P	(*0.000*)		***p***	(*0.000*)	

Table 14 Correlations of Perceived Benefit in experiment 3

Finally, we checked the reliability of the scale we used. A reliable scale should preferably have Cronbach's alpha of at least 0.7. With the number of items per scale and the mean correlation, this value can be calculated. In addition, we checked for all constructs whether the reliability increases when each of the scale items is deleted. Table 15-Table 17 shows the reliabilities of the constructs. The reliability of a scale if an item is deleted is only shown if removing an item increases the reliability.

Data exchange application			
Pre-test		**Post-test**	
Construct	Cronbach's α	Construct	Cronbach's α
CX_E	0.471	CX	0.555
CX if CX_E3 deleted	0.508	CX if CX_E3 deleted	0.701
CX if CX_E3 and CX_E4 deleted	0.588	TW_R	0.623
TW_E	0.546		

Table 15 Reliability of constructs in experiment 1

Table 15 shows the reliabilities of the constructs of the experiment with the data exchange application. The reliability of **Perceived Complexity** increases when items 3 and 4 are removed. This corresponds with our previous findings, namely a low or insignificant correlation with the other items of this scale; this item will thus be removed. The values are not very high, but not such that the constructs cannot be formed. However, these reliabilities should be kept in mind when we interpret results in further analyses.

Preferably, the items for one scale should be kept the same for the constructs in the pre-test and post-test because we need to compare these in further analyses. A dilemma arises: the reliability of **Perceived Complexity** in the pre-test is higher when item 3 and item 4 are removed, while the reliability of **Perceived Complexity** becomes higher when only item 3 is removed. The question is whether item 4 in the pre-test should be included or excluded. We conclude that it is important that the constructs are reliable, as their Cronbach's alpha is relatively low compared to the post-test constructs. Therefore, item 3 and 4 will be removed. For **Perceived Trustworthiness**, deleting an item does not increase the reliability because its scale consists of two items.

De-anonymization application			
Pre-test		**Post-test**	
Construct	Cronbach's α	Construct	Cronbach's α
CX_E	0.546	CX	0.715
CX if CX_E3	0.576	CX if CX_E3 deleted	0.725
CX if CX_E3 & CX_E2 deleted	0.606	TW_R	0.623
TW_E	0.473		

Table 16 Reliability of constructs in experiment 2

Table 16 shows the reliabilities of the constructs of experiment 2 with the de-anonymization application. We conclude that item 3 of **Perceived Complexity** should be excluded from the analysis. Also, the same

dilemma applies to this experiment as in the experiment with the data exchange application: the reliability of the item increases when deleting items 2 and 3 of **Perceived Benefit** instead of only removing item 3. Again, we prefer a reliable scale over a reliable scale, and both items will be removed. For **Perceived Trustworthiness**, removing an item does not increase the reliability since its scale consists of only two items.

Data valuation application			
Pre-test		**Post-test**	
Construct	Cronbach's α	Construct	Cronbach's α
CX_E	0.567	CX	0.730
CX if CX_E3 deleted	0.665	CX if CX_E3 deleted	0.785
BN_E	0.714	BN_R	0.892

Table 17 Reliability of constructs in experiment 3

Table 17 shows the reliabilities of the constructs of the experiment with the data valuation application. We conclude that item 3 of **Perceived Complexity** should be excluded from the analysis in both pre-test and post-test and item 2 from the pre-test because it increases the reliability. The reliabilities in this experiment seem relatively high in comparison to the data exchange and de-anonymization application. For **Perceived Benefit**, removing an item does not increase the reliability since its scale consists of only two items.

Despite the low reliabilities, we formed the constructs. In interpreting further analyses, we should keep in mind that the reliability of **Perceived Complexity** and **Perceived Trustworthiness** is unsatisfactory in the pre-test of the data exchange and the de-anonymization application. To increase the interpretability of the constructs, the sum score of the items is divided by the number of items it consists of. In other words: the average of these items is calculated.

2.2.2 Assumption checks

In this section, we compare the pre-test and post-test factors. These are presented as the expectations and the results, as we measured the expectations for a data exchange application in the pre-test and the actual result of our demonstrators in the post-test. The factors are compared with paired-samples T-tests. We used this test because we have repeated measurements of the same group of respondents; there is a pre-test and a post-test for each factor. It should be noted that item 2 of **Willingness to Use** (WU_R2) only has a post-test item. Therefore, we could not perform a paired-samples T-test. As a result, we used a benchmark value to compare this item. We performed a one-sample T-test with 3 (the midpoint of the Likert scale) as the test value.

The data must meet four assumptions to make sure we can perform a paired-samples t-test. First, the dependent variable must be measured on a continuous scale. The factors of the pre-test and post-test for each experiment are all measured on a 5-point Likert scale that we see as an interval value. Secondly, the samples we compare should consist of related respondents. In other words, the same respondents are present in the pre-test and the post-test. All respondents are measured in the pre-test and in the post-test in our experiments. The groups we compare are thus related. Thirdly, the outliers should be removed from the data. We use a predefined scale, the 5-point Likert scale, and therefore assume that there are no outliers.

Lastly, the dependent variables should be normally distributed. We checked the skewness and kurtosis of the data and performed the Kolmogorov-Smirnov (KS) and the Shapiro-Wilk (SW) test to determine whether the data is normally and symmetrically distributed. Table 18-Table 20 shows the results for each experiment. According to the KS- and SW-tests, the variables are not normally distributed. For each variable, the test value is significant, and therefore we reject the initial hypothesis that states that the data are normally distributed. We can still assume the data are not severally non-normally distributed

when the skewness and kurtosis are near 0. Except for TW_R in experiments 1 and 2 and WU_R2 in experiment 3, the data is moderately to highly skewed. We use 1 as a threshold to determine whether the skewness or kurtosis is too high for the data to be symmetrically distributed; every value lower than -/+ 1 is within an acceptable range. We see that all the variables have a negative skewness, which means it 'looks more like' a uniform distribution. Regarding the kurtosis, most data have a positive value, which means that the distributions of our dependent variables have (sharp) peaks.

Some variables exceeded our threshold for the skewness and the kurtosis and do thus not meet the assumption of normality. However, given the relatively large sample size, we did proceed with the statistical tests.

1. Data exchange application								
Item*	**Skewness**	**Std. Er.**	**Kurtosis**	**Std. Er.**	**KS**	***p***	**SW**	***p***
BN_E	-1.205	0.178	2.499	0.354	0.282	0.000	0.752	0.000
BN_R	-0.718	0.175	0.064	0.347	0.245	0.000	0.860	0.000
CT_E	-1.027	0.174	0.278	0.346	0.267	0.000	0.797	0.000
CT_R	-0.589	0.174	0.016	0.346	0.289	0.000	0.854	0.000
CX_E	-1.391	0.180	2.857	0.358	0.207	0.000	0.833	0.000
CX_R	-1.028	0.174	1.307	0.346	0.158	0.000	0.905	0.000
SC_E	-1.285	0.175	2.102	0.347	0.269	0.000	0.775	0.000
SC_R	-0.641	0.174	0.342	0.346	0.269	0.000	0.858	0.000
TW_E	-1.195	0.176	2.102	0.350	0.205	0.000	0.857	0.000
TW_R	-0.453	0.174	-0.067	0.346	0.208	0.000	0.929	0.000
WU_E	-0.650	0.175	-0.565	0.347	0.241	0.000	0.852	0.000
WU_R	-0.888	0.174	0.750	0.346	0.312	0.000	0.836	0.000
WU_R2	-0.744	0.175	0.657	0.348	0.292	0.000	0.850	0.000

** E stands for Expectation, and are thus pre-test items. R stands for Results, and are thus post-test items*

Table 18 Normality and symmetrically checks for experiment 1

2. De-anonymization application								
Item*	**Skewness**	**Std. Er.**	**Kurtosis**	**Std. Er.**	**KS**	***p***	**SW**	***p***
BN_E	-1.034	0.179	**0.397**	0.356	0.315	0.000	0.758	0.000
BN_R	-0.503	0.179	**-0.497**	0.355	0.232	0.000	0.864	0.000
CT_E	-1.472	0.178	2.611	0.355	0.350	0.000	0.707	0.000
CT_R	-0.746	0.177	0.146	0.353	0.276	0.000	0.859	0.000
CX_E	-1.213	0.178	1.352	0.355	0.206	0.000	0.831	0.000
CX_R	-0.787	0.177	0.516	0.353	0.165	0.000	0.927	0.000
SC_E	-1.292	0.178	1.324	0.355	0.368	0.000	0.695	0.000
SC_R	-0.776	0.178	0.404	0.354	0.284	0.000	0.850	0.000
TW_E	-1.244	0.178	1.837	0.355	0.233	0.000	0.804	0.000
TW_R	-0.470	0.177	-0.070	0.353	0.226	0.000	0.917	0.000
WU_E	-1.254	0.179	1.594	0.355	0.257	0.000	0.768	0.000
WU_R	-1.017	0.177	0.770	0.352	0.311	0.000	0.816	0.000
WU_R2	-0.711	0.178	0.011	0.354	0.294	0.000	0.858	0.000

* E stands for Expectation, and are thus pre-test items. R stands for Results, and are thus post-test items

Table 19 Normality and symmetrically checks for experiment 2

Data valuation application								
Item*	**Skewness**	**Std. Er.**	**Kurtosis**	**Std. Er.**	**KS**	***p***	**SW**	***p***
BN_E	-2.241	0.179	8,440	0.356	0.268	0.000	0.721	0.000
BN_R	-0.741	0.179	0.047	0.355	0.195	0.000	0.915	0.000
CX_E	-1.500	0.182	3.979	0.361	0.180	0.000	0.837	0.000
CX_R	-1.053	0.178	1.017	0.354	0.209	0.000	0.899	0.000
TW_E	-1.445	0.179	3.287	0.355	0.306	0.000	0.717	0.000
TW_R	-0.999	0.178	1.385	0.355	0.343	0.000	0.798	0.000
WU_E	-0.859	0.178	0.452	0.354	0.251	0.000	0.824	0.000
WU_R	-0.785	0.178	0.376	0.355	0.287	0.000	0.856	0.000
WU_R2	**-0.435**	0.178	-0.119	0.354	0.282	0.000	0.880	0.000

* E stands for Expectation, and are thus pre-test items. R stands for Results, and are thus post-test items

Table 20 Normality and symmetrically checks for experiment 3

2.2.3 Comparing the pre-test and post-test of experiment 1

Table 21 shows the descriptive statistics of the factors in experiment 1. The scales used range from 1 (Totally disagree) to 5 (Totally agree). The means of the factors in the post-test are all greater than 3, indicating that the perception of our data exchange application is predominantly positive because its values are greater than the midpoint of the Likert scale (3 = neither agree nor disagree). We also found that the perception of the demonstrator is by no means higher evaluated than the expectations of such an application. We expected this to be the case, as the pre-test measures an ideal data exchange application that meets all the respondents' requirements. It is likely that the actual application does not meet all these requirements. The question is then how much the perception differs from the expectations, so we know the areas of improvement concerning the application. Further analyses must show whether these differences in means are statistically significant.

Factor	Pre-/post-test	N	Mean	Median	Std. Dev.
Perceived Benefit	Expectation	187	4.364	4.000	0.708
	Result	194	3.910	4.000	0.970
Perceived Control	Expectation	195	4.150	4.000	0.988
	Result	196	3.810	4.000	0.877
Perceived Complexity	Expectation	182	4.346	4.500	0.668
	Result	196	4.103	4.333	0.681
Perceived Security	Expectation	191	4.290	4.500	0.815
	Result	195	3.850	4.000	0.873
Perceived Trustworthiness	Expectation	194	4.358	4.000	0.637
	Result	196	3.905	4.000	0.689
Willingness to Use	Expectation	194	3.890	4.000	1.057
	Result	196	3.850	4.000	0.902

Table 21 Descriptive statistics of factors in experiment 1

In the next analyses, we statistically compared the factors of the pre-test and post-test. Table 22 and Figure 2 present the results of the paired-samples T-test for experiment 1. We found a statistical difference for **Perceived Benefit** [t(2,184)=5.051, p=0.000], **Perceived Control** [t(2,194)= 3.778, p=0.000], **Perceived Complexity** [t(2,181)= 3.657, p=0.000], **Perceived Security** [t(2,193)= 5.362, p=0.000], **Perceived Trustworthiness** [t(2,189)= 6,877, p=0.000] at the p<0.05 level. For these factors, the expectation of the application differs from the actual perception.

Factor	Df	Mean difference	Std. Dev.	t	*p*
Perceived Benefit	184	0.427	1.150	5.051	0.000
Perceived Control	194	0.338	1.251	3.778	0.000
Perceived Complexity	181	0.244	0.899	3.657	0.000
Perceived Security	193	0.443	1.151	5.362	0.000
Perceived Trustworthiness	189	0.453	0.907	6,877	0.000
Willingness to Use	193	0.041	1.250	0.459	0.646

Table 22 Significance of mean differences of factors in experiment 1

We see multiple explanations that could explain this difference. Firstly, the respondents got confused by the results of the tasks they had to perform: some respondents thought they should be able to find the zip codes in the list, even though it was not. This could have spread confusion. Also, they found the process too complex and did not understand the essence of privacy-preserving technology. Not every respondent may know how reliable the application actually was. Lastly, participants mentioned that they found the application insufficiently transparent.

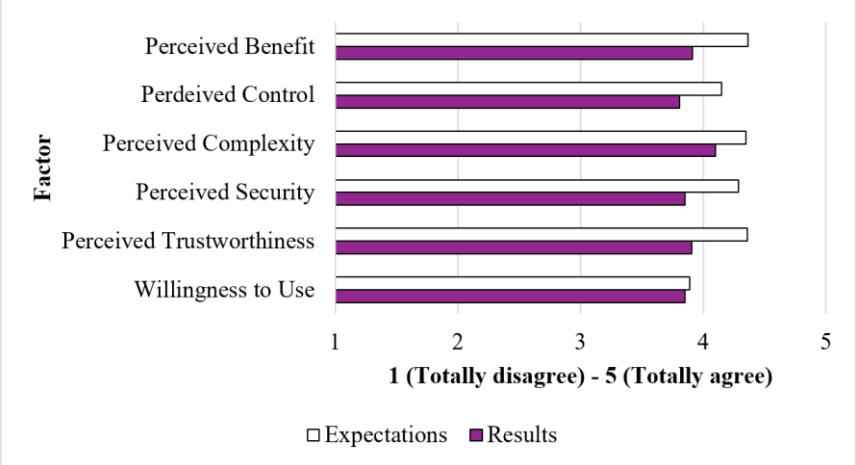


Figure 2 Mean differences between expectations and results in experiment 1

The difference for **Perceived Benefit** differs the most from the results. A commonly mentioned reason for unwillingness to use was that respondents thought that the application would not be beneficial in their type of job. The reason for the relatively big difference could thus be the applicability of the application.

Factor	Descriptive statistics				One-sample T-test			
	N	Mean	Std. Dev.	Median	Mean difference	Test value	t	*p*

Willingness to Use: I would be willing to use the application in the next 6 months (WU_R2)	193	3.684	1.012	4	0.684	3.000	9.252	0.000

Table 23 Benchmark of Willingness to Use in experiment 1

Willingness to Use [t(2,193)= 0.459, p=0.646] is not statistically significant at this level. This can mean that the actual perception lives up to the expectation for such an application or that the respondents expect not to use it either way. Based on the one-sample T-test (see Table 23), **Willingness to Use** [t(192)=9.252, p=0.000] is greater than the benchmark value of three. This indicates that there is an actual **Willingness to Use** the data exchange application. Research of the open question concerning the reasons of use showed that the advantages of the application are the ease of use and the possibility to share confidential data with a smaller risk of data leaks.

2.2.4 Comparing the pre-test and post-test of experiment 2

Table 24 presents the descriptive statistics of the factors in experiment 2. The means of the factors in the post-test are all greater than the midpoint of the Likert scale, which indicates that the respondents are predominantly willing to use de-anonymization application.

As in experiment 1, the actual perception of the demonstrator is by no means higher evaluated than the expectations of such an application. Further analyses must show whether these mean differences are statistically significant.

Factor	Pre-test/post-test	N	Mean	Median	Std. Dev.
Perceived Benefit	Expectation	184	4.380	5.000	0.749
	Result	185	3.880	4.000	0.961
Perceived Control	Expectation	186	4.460	5.000	0.737
	Result	188	3.790	4.000	0.990
Perceived Complexity	Expectation	186	4.381	4.500	0.667
	Result	188	3.996	4.000	0.752
Perceived Security	Expectation	186	4.520	5.000	0.669
	Result	187	3.860	4.000	0.934
Perceived Trustworthiness	Expectation	186	4.541	4.500	0.497
	Result	188	3.960	4.000	0.711
Willingness to Use	Expectation	185	4.280	4.000	0.832
	Result	189	3.920	4.000	0.977

Table 24 Descriptive statistics of factors in experiment 2

Next, we compared the factors of the pre- and post-tests from experiment 2 with a paired-samples T test (see Table 25 and Figure 3). We found a statistical mean difference at $p<0.05$ level for all factors: **Perceived Benefit** [t(179)=6.499, p = 0.000], **Perceived Control** [t(184)=7.619, p = 0.000**], Perceived Complexity** [t(184)=5.564, p = 0.000], **Perceived Security** [t(183)=8.002, p = 0.000], **Perceived Trustworthiness** [t(184)=9.796, p = 0.000] and **Willingness to Use** [t(184)=4.477, p = 0.000].

Factor	Df	Mean difference	Std. Dev.	t	*p*
Perceived Benefit	179	0.506	1.044	6.499	0.000
Perceived Control	184	0.665	1.187	7.619	0.000
Perceived Complexity	184	0.385	0.940	5.564	0.000

Perceived Security	183	0.658	1.115	8.002	0.000
Perceived Trustworthiness	184	0.581	0.807	9.796	0.000
Willingness to Use	184	0.362	1.100	4.477	0.000

Table 25 Significance of mean differences of factors in experiment 2

Perceived Control, **Perceived Security**, and **Perceived Trustworthiness** have the highest mean differences. We can explain this using the open questions regarding the unwillingness to use the de-anonymization application. These answers show concern about the reliability of the application and what this means for their privacy and security. Responses indicate that participants do not fully understand the process of the de-anonymization application and hence do not fully understand how their privacy and security are guaranteed. Also, responses show distrust in the application provider because they are afraid the provider misuses the data by using or selling it.

As mentioned, not all respondents do fully understand how the process works. They find the process confusing and difficult. This explains the mean difference of **Perceived Complexity**. Meanwhile, the significant mean difference of **Perceived Benefit** can be attributed to the same reason as in experiment 1, namely the irrelevance of the de-anonymization application at the respondents' current job.

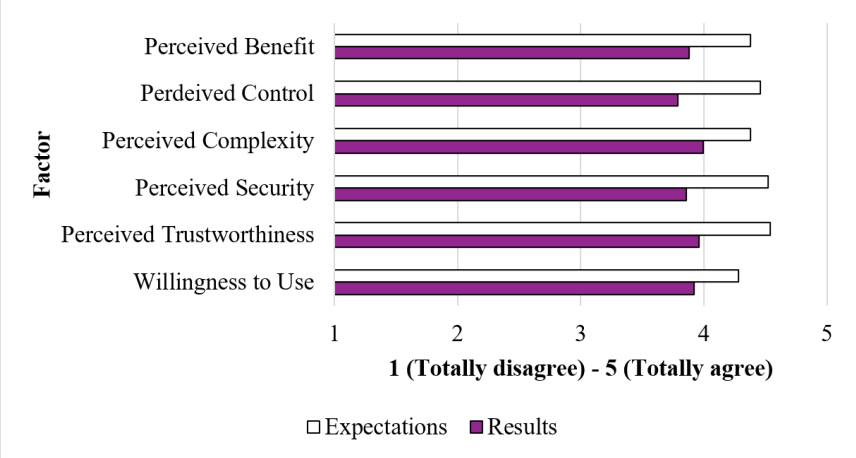


Figure 3 Mean differences between expectations and results in experiment 2

We performed a One-sample T-test for the second item of **Willingness to Use** (see Table 26). **Willingness to Use** [t(186)=11.553, p=0.000] is greater than the benchmark value of three. We can conclude that there exists an actual **Willingness to Use** the de-anonymization application. From an open question, respondents indicated that de-anonymization makes them feel secure in protecting sensitive data and may help their companies be GDPR-compliant. However, not all respondents share this opinion, as we mentioned in the previous paragraph. The lack of understanding of what the application does seems to be one of the causes.

Factor	Descriptive statistics				One-sample T-test			
	N	Mean	Std. Dev.	Median	Mean difference	Test value	t	*p*
Willingness to Use: I would be willing to use the application in the next 6 months (WU_R2)	187	3.746	0.897	4.000	0.746	3.000	11.553	0.000

Table 26 Benchmark of Willingness to Use in experiment 2

2.2.5 Comparing the pre-test and post-test of experiment 3

Table 27 presents the descriptive statistics of the factors in experiment 3. The means of the factors in the post-test are all greater than the midpoint of the Likert scale, meaning that the respondents are predominantly willing to use the data valuation application. As in experiment 1, the actual perception of the demonstrator is by no means higher evaluated than the expectations of such an application. Further analyses must show whether these mean differences are statistically significant.

Factor	Pre-test/post-test	N	Mean	Median	Std. Dev.
Perceived Benefit	Expectation	184	4.566	4.750	0.589
	Result	185	3.731	4.000	0.967
Perceived Complexity	Expectation	179	4.441	4.667	0.592
	Result	187	3.909	4.000	0.837
Perceived Trustworthiness	Expectation	185	4.430	5.000	0.690
	Result	186	3.860	4.000	0.854
Willingness to Use	Expectation	187	4.110	4.000	0.869
	Result	186	3.760	4.000	0.974

Table 27 Descriptive statistics of factors in experiment 3

Next, we compared the factors of the pre- and post-tests from experiment 3 with paired-samples T tests (see Table 28 and Figure 4). We found the following significant mean differences at $p<0.05$ level: **Perceived Benefit** [$t(181)=10.730$, $p = 0.000$], **Perceived Complexity** [$t(178)=7.691$, $p = 0.000$], **Perceived Trustworthiness** [$t(183)=7.580$, $p = 0.000$] and **Willingness to Use** [$t(185)=4.265$, $p = 0.000$].

Factor	Df	Mean difference	Std. Dev.	t	*p*
Perceived Benefit	181	0.835	1.050	10.730	0.000
Perceived Complexity	178	0.533	0.926	7,691	0.000
Perceived Trustworthiness	183	0.565	1.011	7,580	0.000
Willingness to Use	185	0.344	1.100	4.265	0.000

Table 28 Significance of mean differences of factors in experiment 3

Perceived Benefit has a relatively high mean difference (0.835). In the open questions about unwillingness to use, a dozen respondents indicated that the data valuation application is neither beneficial nor relevant for their work and therefore not useful. An explanation for the difference could thus be that the possibilities of the application are unclear or that there is just no demand from the individual consumer. Meanwhile, concerning the significant mean difference of **Perceived Complexity**, respondents suggested that the application is too complex for their abilities. The many and complicated options and choices that have to be made manually seem to cause this concern, resulting in a time-consuming process.

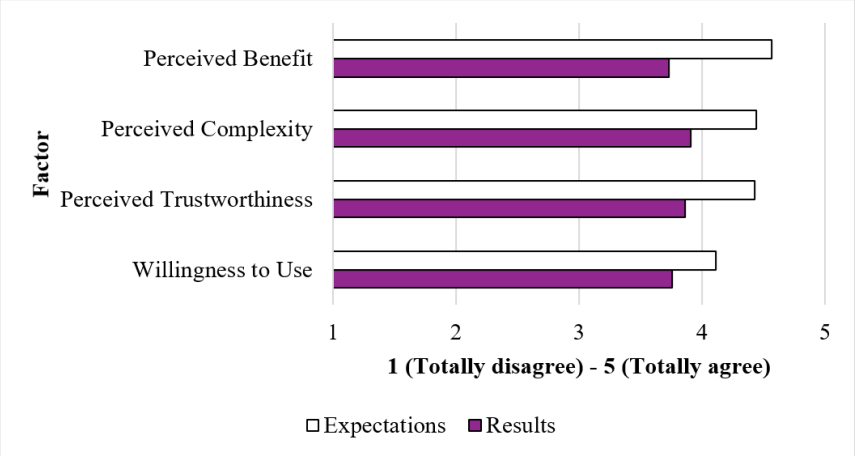


Figure 4 Mean differences between expectations and results in experiment 3

We then performed a One-sample T-test for the second item of **Willingness to Use** (see Table 29) and found that it is greater than the benchmark value of three [t(186)=9.252, p=0,000]. We can conclude that there exists an actual **Willingness to Use** the data valuation application. Respondents pointed out from an open question that the data valuation application can provide valuable and detailed insights into their data.

Factor	Descriptive statistics				One-sample T-test			
	N	Mean	Std. Dev.	Median	Mean difference	Test value	t	*p*
Willingness to Use: I would be willing to use the application in the next 6 months (WU_R2)	187	3.684	1.012	4.000	0.684	3.000	9.252	0.000

Table 29 Benchmark of Willingness to Use in experiment 3

2.2.6 Alternative statistical tests

Now that we analyzed the three applications separately, it is interesting to compare the **Willingness to Use** each application. Besides meeting the assumptions mentioned in section 2.2.2, the variances of the dependant variable need to be equal in the three groups. To do this, we performed Levene's test of equal variances. The variances are equal at a significance level of 0.05 [F(2,564)=2.613, p =0.074]. Therefore, we should perform a One-way ANOVA. This test shows that the **Willingness to Use** does not differ between the applications [F(2,564)=0.336, p=0.715]. In other words, there is no significant difference between the demonstrators in terms of how much respondents are willing to use them.

Next, we made three regression models to predict **Willingness to Use** (WU_R2), as stated in Table 30. In the questions with open answers about (un)**Willingness to Use**, reasons such as relevance, complexity, and trust in the application were mentioned several times. Therefore, we have chosen Perceived Benefit, **Perceived Complexity**, **Perceived Trustworthiness**, and **Willingness to Use** (WU_R1) as predictors for all experiments and added **Perceived Control** and **Perceived Security** for experiments 1 and 2. The post-test results are included because we want to measure how the perceptions of our demonstrators may influence the use of it.

Before interpreting the results, we need to assess whether there is multicollinearity. To do this, the variance inflation factor (VIF) was requested for the predictors in the models. As the VIF increases, the interrelationships between the predictors are stronger. In general, it is stated that a VIF value of 4 or higher indicates there is multicollinearity (Gordon, 2015). The VIF of the predictors in the regression

models is less than 3 (see Table 30) and does thus not exceed this limit. It is therefore assumed that multicollinearity is not a problem.

Experiment	1. Data exchange application WU_R2*			2. De-anonymization application WU_R2**			3. Data valuation application WU_R2***		
Factor	**β**	***p***	**VIF**	**β**	***p***	**VIF**	**β**	***p***	**VIF**
Perceived Benefit	**0.499**	**0.000**	1.388	**0.333**	**0.000**	1.242	**0.563**	**0.000**	1.683
Perceived Control	**0.188**	**0.012**	1.930	-0.009	0.908	1.899			
Perceived Complexity	**0.157**	**0.022**	1.654	**0.235**	**0.003**	1.763	**0.276**	**0.001**	2.859
Perceived Security	0.051	0.512	2.126	**0.172**	**0.047**	2.174			
Perceived Trustworthiness	0.003	0.965	1.667	0.079	0.912	2.212	-0.047	0.564	2.588

**F(5,170)=36.961, p = 0.000, R^2=0.521*

***F(5,177)=23.234, p = 0.000, R^2=0.396*

****F(3, 180)=70.071, p = 0.000, R^2=0.539*

Table 30 Regression model of Willingness to Use in all applications

Table 30 presents the three regression models. The R^2 of the regression models is relatively high: the regression models explain 52.1% (model 1), 39.6% (model 2), and 53.9% (model 3) of the variance of **Willingness to Use**. This indicates that these models fit the collected data. Furthermore, we found several interesting significant predictors.

Firstly, **Perceived Benefit** is in all models significant at a significance level of >0.05. When **Perceived Benefit** increases by one standard deviation, the **Willingness to Use** increases by 0.499 in experiment 1, 0.333 in experiment 2, and 0.563 in experiment 3. We can thus conclude that if users perceive the application as beneficial for their company or job, this can contribute to their **Willingness to Use** the application.

Secondly, **Perceived Complexity** is significant at a significance level of >0.05 in all three regression models. When **Perceived Complexity** increases by one standard deviation, the **Willingness to Use** increases by 0.157 in experiment 1, 0.235 in experiment 2, and 0.276 in experiment 3. Users are more willing to use the applications if the application is clear.

Thirdly, **Perceived Control** is significant in experiment 1 with a significance level of 0.05. When **Perceived Complexity** increases by one standard deviation, the **Willingness to Use** increases by 0.159. When the data exchange applications make the user feel more in control over their data, this contributes to the **Willingness to Use**. **Perceived Security** is significant in experiment 2 at a significance level of 0.05. When **Perceived Security** increases by one standard deviation, the **Willingness to Use** increases by 0.172. Participants that feel more secure in using the de-anonymization application are more willing to use it. It is noticeable that **Perceived Trustworthiness** is in none of the models a significant factor. This contradicts existing research. Pavlou (2003), for example, found that trust in e-commerce is a direct antecedent of the intention to transact online. An explanation could be the low reliability of the **Perceived Trustworthiness** scale (see section 2.2.1)

2.2.7 Testing of control variables

In addition to the regression models, we also analyzed **Willingness to Use** based on different demographic groups, namely (1) data-related role; (2) involvement in developing new products; and (3) familiarity with technology in focus (MPC/PSI, de-anonymization, or data valuation). For every experiment, we split the respondents into two groups for these demographics. The mean differences of

the **Willingness to Use** between these groups are compared. Specifically, the respondents that are never, rarely, or sometimes involved in developing new products are assigned to one group ('not involved'), and the respondents that are often or always involved in developing new products are assigned to the second group ('involved'). Regarding the data-related role at work, respondents that are not or slightly involved in a data-related role are assigned to one group ('not related'), and the people that are somewhat, moderately, or very involved are assigned to the second group ('related'). Lastly, the respondents who are not or slightly familiar with the application are assigned to one group ('not familiar'), and the respondents who are somewhat, moderately, or very familiar with the application are assigned to the second group ('familiar').

Before comparing the means, we checked whether the dependent factor is normally distributed. We checked the skewness and kurtosis of the data and performed the Kolmogorov-Smirnov (KS) and Shapiro-Wilk (SW) test to determine whether the data is normally and/or symmetrically distributed (Table 31). According to the KS- and SW-tests, the variables are not normally distributed ($p<0.05$). The skewness and kurtosis values are all lower than |1.000|, which means that the data is not too moderately skewed and curved. With the large sample size in mind, it can be concluded that we can proceed with the independent samples T-test.

#	Item	Group	N	Skewness	Std. Er.	Kurtosis	Std. Er.	KS	*p*	SW	*p*
1. Data exchange	Data-related role	Not related	119	-0.802	0.224	0.596	0.444	0.296	0.000	0.853	0.000
		Related	77	-0.384	0.276	-0.208	0.545	0.270	0.000	0.846	0.000
	Involvement in developing new products	Not involved	129	-0.737	0.311	0.215	0.613	0.275	0.000	0.871	0.000
		Involved	67	-0.592	0.209	0.510	0.416	0.292	0.000	0.848	0.000
	Familiarity with MPC/PSI	Not familiar	156	-0.773	0.196	0.834	0.390	0.293	0.000	0.850	0.000
		Familiar	40	-0.824	0.374	0.299	0.733	0.277	0.000	0.820	0.000
2. De-anonymization	Data-related role	Not related	104	-0.733	0.239	0.278	0.474	0.280	0.000	0.867	0.000
		Related	82	-0.783	0.266	-0.020	0.526	0.302	0.000	0.844	0.000
	Involvement in developing new products	Not involved	128	-0.733	0.239	0.278	0.474	0.224	0.000	0.898	0.000
		Involved	61	-0.783	0.266	-0.020	0.526	0.320	0.000	0.830	0.000
	Familiarity with de-anonymization	Not familiar	126	-0.419	0.314	-0.174	0.618	0.287	0.000	0.860	0.000
		Familiar	60	-0.867	0.213	0.277	0.423	0.294	0.000	0.854	0.000
3. Data valuation	Data-related role	Not related	109	-0.204	0.231	0.090	0.459	0.283	0.000	0.880	0.000
		Related	78	-0.743	0.272	-0.118	0.538	0.279	0.000	0.857	0.000
	Involvement in developing new products	Not involved	122	-0.370	0.219	-0.210	0.435	0.298	0.000	0.870	0.000
		Involved	65	-0.582	0.297	0.158	0.586	0.248	0.000	0.899	0.000
	Familiarity with data valuation	Not familiar	84	-0.070	0.263	-0.401	0.520	0.252	0.000	0.899	0.000
		Familiar	103	-0.803	0.238	0.507	0.472	0.299	0.000	0.849	0.000

Table 31 Normality checks for Willingness to Use grouped by social demographics

Table 32 presents the results of the independent T-tests. Regarding the data exchange application, we found a significant difference in **Willingness to Use** between the respondents that (do not) have a data-related role at work [t(2,191)=-2,468, p=0.015] and the respondents that are (not) familiar with MPC or PSI [t(2,191)=-2,437, p=0.016] at a significance level of <0.05. There is a significant difference at a level of <0.100 between the groups that are (not) involved in developing new products. Users that have a data-related role at work, are involved in developing new products, and are familiar with MPC are

more willing to use the data exchange application. We found no significant differences for experiment 2. Regarding the data valuation application, we found that respondents familiar with data valuation are more willing to use the application [t(2,185)= -2,178, p=0.032].

Data exchange application								
Factor	Group	Descriptive statistics			Independent samples T-test			
		N	Mean	Std. Dev.	Df*	Mean difference	t	*p (2-tailed)*
Data related role	Not related	117	3.624	0.944	191	-0.310	-2,468	0.015
	Related	76	3.934	0.789				
Involvement in developing new products	Not involved	59	3.559	1.038	191	-0.269	-1.764	0.081
	Involved	134	3.828	0.818				
Familiarity with MPC/PSI	Not familiar	156	3.667	0.889	191	-0.383	-2,437	0.016
	Familiar	40	4,050	0.879				
De-anonymization application								
Factor	**Group**	**Descriptive statistics**			**Independent samples T-test**			
		N	**Mean**	**Std. Dev.**	**Df***	**Mean difference**	**t**	***p (2-tailed)***
Data related role	Not related	102	3.608	0.997	191	-0.148	-0.985	0.326
	Related	82	3.756	1.037				
Involvement in developing new products	Not involved	58	3.500	1.013	185	-0.244	-1.520	0.130
	Involved	129	3.744	1.018				
Familiarity with de-anonymization	Not familiar	124	3.653	1.012	182	-0.063	-0.397	0.692
	Familiar	60	3.717	1.027				
Data valuation application								
Factor	**Group**	**Descriptive statistics**			**Independent samples T-test**			
		N	**Mean**	**Std. Dev.**	**Df***	**Mean difference**	**t**	***p (2-tailed)***
Data related role	Not related	109	3.624	0.959	185	-0.145	-0.969	0.334
	Related	78	3.769	1.231				
Involvement in developing new products	Not involved	122	3.639	0.988	185	-0.130	-0.835	0.405
	Involved	65	3.769	1.057				
Familiarity with data valuation	Not familiar	138	3.594	1.030	185	-0.345	-2,178	0.032
	Familiar	49	3.939	0.922				

Table 32 Comparison of mean difference of Willingness to Use grouped by social demographics

2.3 Conclusions

This study examined users' perceptions of the data exchange, de-anonymization, and data valuation application developed in Safe-DEED. Specifically, we have studied the Perceived Benefit, complexity, control, security, trustworthiness, and Willingness to Use. We studied the difference between the expectations and the actual perception of our applications.

In the PAF analysis, we found that the reliability of the constructs Perceived Complexity and Perceived Trustworthiness is unsatisfactory in the pre-test of the data exchange and the de-anonymization application. We should consider this when interpreting the results of these factors.

Secondly, we found that the applications are perceived predominantly positive. However, the actual perception of the demonstrators did not meet the expectations for such applications. Among other things, respondents mentioned the complex processes, relevance, and security concerns as reasons for this. Nevertheless, these constructs did score higher than the middle of the scale for all demonstrators, suggesting that Perceived Complexity, Perceived Benefits, Perceived Trustworthiness, and Perceived Security were sufficient.

Next, we further analyzed the Willingness to Use the applications. We found that users are willing to use the applications in the next six months. The Willingness to Use does not differ between the three applications. We found that for all the applications, Perceived Benefit contributes to the Willingness to Use. This indicates that clarifying the relevance and applicability of the application and how to benefit from it may result in a greater Willingness to Use it. Also, users are more willing to use the applications when they find the application clear and easy to use. Users who perceive the applications as less complex are more willing to use the application. From this, it follows that the application could be improved by including more and clearer explanations to reduce the Perceived Complexity. Regarding the data exchange application, we found that users who feel in control over their data are more willing to use the application. In addition, we found that participants that feel more secure in using the de-anonymization application are more willing to use it. It is remarkable that, in the other experiments, Perceived Security or Perceived Trustworthiness were not significant antecedents of Willingness to Use the applications. On the one hand, this is well explainable because security is often an afterthought and not a decisive factor in using a service. Another explanation could be that the context of the demonstrator may not feel threatening to the users since it is not their (companies) data. Therefore, security and trust may feel less important to them.

Lastly, we found that users that have a data-related role at work, are involved in developing new products, and are familiar are more willing to use the demonstrator. Also, the users familiar with data valuation are more willing to use the application than users who are not familiar with this technology. For users who are less experienced with privacy-preserving technologies, it may be harder to understand its benefits, how secure it is, and whether they should trust it. The people who are knowledgeable about these technologies may be able to understand the true value of the demonstrators and what a contribution they can make to data sharing between companies. On the one hand, this information identifies a target audience that is willing to use the application. On the other hand, it highlights possibilities to improve the application to appeal to the other group: people unfamiliar with the technology. For example, the inner working of the technology and the potential user benefit could be explained more.

3. Study 2: Comparison to trusted-third party scenario

The first study provided insights into how the Safe-DEED technologies positively affect business users' perception of trust and intention to use within specific use-cases developed in WP4, WP5, and WP6. However, whether it could outperform existing approaches to data sharing that are dominant in today's data economy remains unclear. Therefore, we conducted a follow-up study (study 2) to evaluate the Safe-DEED technologies in another setting, namely the emerging context of personal data marketplaces. While we do not evaluate the Safe-DEED demonstrators per se, we do compare the underlying notions of decentralized data collaboration through MPC with traditional approaches of trusted third parties that collect and distribute data.

To scope our study, we focus only on MPC as one part of the technology developed in Safe-DEED. Specifically, we investigate whether MPC induces the feeling of control over data and trust towards other parties (i.e., data marketplaces operator and data buyers) among individual citizens. Compared to study 1, this study (1) focuses on consumers/citizens rather than businesses; (2) compares the Safe-DEED types of technologies of `trustless' / decentralized data collaboration to the current standard of trusted third parties. We also seek to understand if MPC could reduce risk perception and Privacy Concerns in data sharing. Ultimately, we examine the impact of MPC on individuals' Willingness to Share their data through data marketplaces. Prior studies have suggested the importance of these factors in individual data sharing decisions, which might hinder the realization of the data economy (see Table 33). Thus, MPC can be seen as an enabler in maintaining sufficient control in data sharing without harming individuals' privacy.

Concept	Definition	Selected relevant studies
Perceived Control	An individual's beliefs in his or her ability to manage the release and dissemination of personal information.	Brandimarte et al. (2013); Dinev et al. (2013); Hajli & Lin (2016); Krasnova et al. (2010); Spiekermann (2005); Xu et al. (2011)
Perceived Risk	The expectation of losses associated with the disclosure of personal information.	Dinev et al (2013); Kehr et al. (2015); Malhotra et al. (2004); Pavlou (2003); Xu et al. (2011)
Privacy Concerns	Beliefs about who has access to information that is disclosed when using the Internet and how it is used.	Derikx et al. (2016); Dinev & Hart (2006); Kato et al. (2016); Kehr et al. (2015); Malhotra et al. (2004)
Trust	An individual's confidence that the data-requesting medium (i.e., data marketplaces, data buyers) will not misuse his/her data.	Dinev & Hart (2006); Kehr et al. (2015); Krasnova et al., (2010); Liu et al. (2005); Malhotra et al. (2004);
Willingness to Share data	An individual's intent to engage in an online exchange relationship with data marketplaces.	Dinev & Hart (2006); Malhotra et al. (2004); Pavlou (2003); Kehr et al. (2015); Krasnova et al., (2010)

Table 33 Definition of factors included in study 2

To fulfil our objective, we compare three different data sharing scenarios involving Trusted Third Party (TTP), MPC, and made-up privacy technology (referred to as Data-Computation-Protection/DCP). In particular, we want to see if MPC performs better than TTP and DCP in terms of our concepts of interest mentioned before. The reason for including a made-up technology in our study is because we use a description of MPC in our experiment rather than a working demonstrator or prototype. A critique could be that users contribute value to the term of MPC rather than to the underlying ideas in the technology. Therefore, we want to see if different privacy technologies would make any differences in perception or it does not matter for users, even if the technology does not exist. We focus on the specific context of sharing driving data in connected cars as digitalization opens up novel opportunities for value creation from data-driven services (Athanasopolou et al., 2019; Kaiser et al., 2021). However, the sensitive nature of the data generated in this domain has resulted in mounting concerns regarding trust, privacy, risk of data sharing, and control over data (e.g., Docherty et al., 2018). This unique setting makes it interesting to see if the dynamics in data sharing would change with MPC in place.

As this study is still a work in progress, we only present the preliminary results based on the pre-study conducted to test the predefined measurement model. We reported two main findings of this study: the results of the Confirmatory Factor Analysis (CFA) and a one-way ANOVA to compare the means of the three conditions.

3.1 Method

3.1.1 Experimental design

We conducted a controlled, survey-based online experiment to investigate the effect of MPC on the Willingness to Share data in privacy-preserving data marketplaces. We opted for a between-subject design with three experimental conditions (TTP, MPC, and DCP). Each condition is different in terms of the description of the technology, how it works in the data marketplaces, and a screenshot preview of the mock-up (including a disclaimer on how the technology works, see Figure 5). The main difference is that, in TTP, there is a data flow to the central system in which the data will be analyzed and stored there. Meanwhile, in MPC and DCP scenarios, the data is encrypted on the car, and only the analysis results are revealed to the prospective buyer. Both MPC and DCP scenarios are identical, with the name of the technology being the only difference. We did this to see whether the name or the description of the technology matters for participants, even if it is a made-up technology.

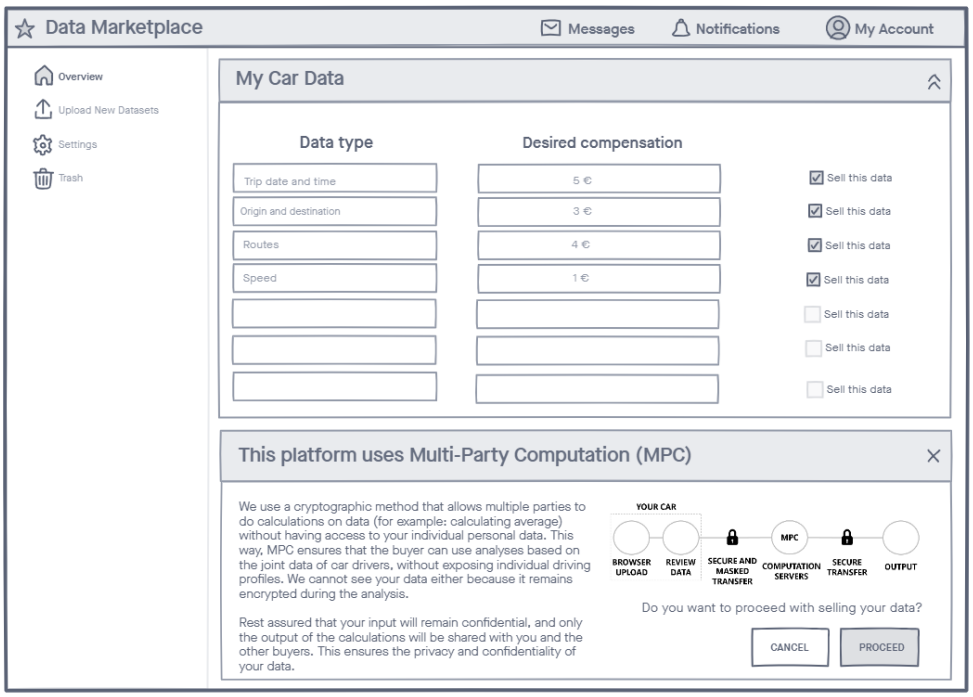


Figure 5 A screenshot preview of the mock-up for the MPC scenario

The experiment consisted of four parts (see Figure 6). After introducing the purpose of the study and the consent form, we presented participants with a persona where they owned a connected car that generated driving data and could sell it via data marketplaces. We asked participants to imagine if mobility service providers are interested in buying their driving data (e.g., trip date and time, destination and routes history, and driving speed) via data marketplaces. Next, we randomly assigned participants to one of the three conditions (TTP, MPC, or DCP) and introduced them to their respective scenarios. We then asked participants to fill in the completion code as proof that they had read and understood the scenario. Subsequently, participants filed out the post-test questionnaire to rate their perception of the data marketplaces presented to them. We concluded the experiment with participants filling out the demographic questions, which are the same for all conditions.

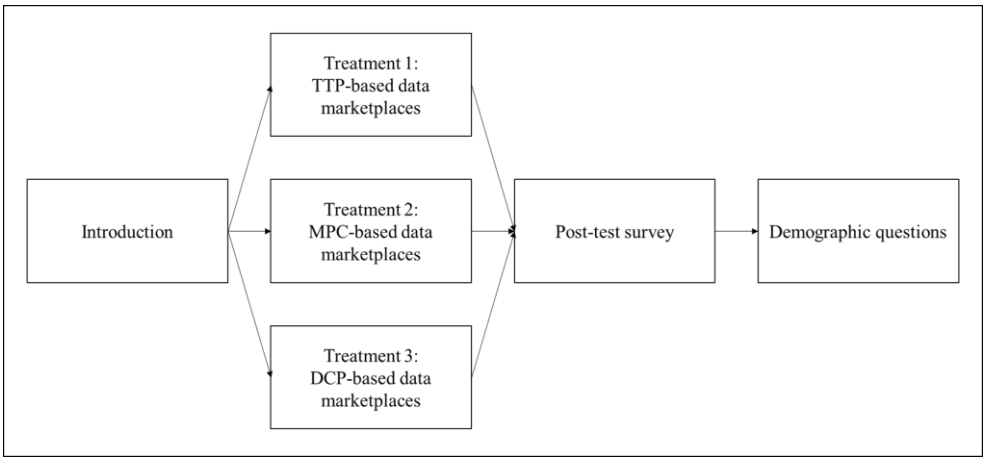


Figure 6 Experimental design overview

3.1.2 Measures

As quantitative studies on sharing driving data via privacy-preserving data marketplaces are lacking, we developed a 5-scale Likert questionnaire based on existing measures used in previous studies in information privacy and e-commerce to fit our context. We modified survey items by Xu, Dinev, Smith, and Hart (2011) to measure both **Perceived Control** and **Perceived Risk**. For **Privacy Concerns**, we adopted measures developed by Dinev and Hart (2006). Meanwhile, for both **Trust in data marketplaces** and **Trust in data buyers**, we used measures by Kehr, Kowatsch, Wenzel, and Fleisch (2015) and adjusted the items based on the actors in question. Finally, to measure **Willingness to Share data**, we used measures by Pavlou (2003). Table 34 presents the final measures.

Additionally, we also asked three privacy-related questions to participants based on the numerous studies conducted by Westin, as summarized by Kumaraguru & Cranor (2005). Depending on the answers (1 = strongly disagree, 4 = strongly agree), we classify participants as privacy-concerned and not privacy-concerned. Then, we further categorized participants into three groups (also known as Westin's Privacy Segmentation Index):

1. **Privacy fundamentalists:** Consumers that are the most protective of their privacy. They feel companies should not be able to acquire personal information for their organizational needs and think that individuals should be proactive in refusing to provide information. Privacy Fundamentalists also support stronger laws to safeguard an individual's privacy.
2. **Privacy unconcerned:** Consumers that are the least protective of their privacy. They feel that the benefits they may receive from companies after providing information far outweigh the potential abuses of this information. Further, they do not favor expanded regulation to protect privacy.
3. **Privacy pragmatists:** Consumers who weigh the potential pros and cons of sharing information; evaluate the protections that are in place and their trust in the company or organization. After this, they decide whether it makes sense for them to share their personal information.

We will use these categories to explore differences among those groups regarding **Perceived Control**, **risk**, **trust**, and **Willingness to Share data**.

Construct	Item	Item wording	Source
Westin's Privacy Segmentation Index	PI1	Consumers have lost all control over how personal information is collected and used by companies.	Kumaraguru & Cranor (2005)
	PI2	Most businesses handle the personal information they collect about consumers in a proper and confidential way.	
	PI3	Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.	
Perceived Control	CTRL_1	I believe I have control over who can access the sensitive data I provided to this data marketplace.	Xu, Dinev, Smith & Hart (2011)
	CTRL_2	I think I have control over what kind of sensitive data is shared by this data marketplace to other companies.	
	CTRL_3	I believe I have control over how other companies use the sensitive data I provided to this data marketplace.	
	CTRL_4	I believe I can control what kind of sensitive data that I can provide to this data marketplace.	
Perceived Risk	RISK_1	I find it risky to provide my sensitive data via this data marketplace.	Xu, Dinev, Smith & Hart (2011)
	RISK_2	I think there is a good chance that my sensitive data will get lost when I provide them via this data marketplace.	

	RISK_3	There would be too much uncertainty associated with providing my sensitive data to this data marketplace.	
	RISK_4	Providing this data marketplace with my sensitive data would involve many unexpected problems.	
Privacy Concerns	PRIV_1	I am concerned that the sensitive data I provide to this data marketplace could be misused.	Dinev & Hart (2006)
	PRIV_2	I am concerned that other parties could find sensitive information about me on this data marketplace.	
	PRIV_3	I am concerned about providing my sensitive data to this data marketplace because of what other parties might do with it.	
	PRIV_4	I am concerned about providing my sensitive data to this data marketplace because it could be used in a way I did not foresee.	
Trust in data marketplaces	TRSD_1	I expect this data marketplace would be trustworthy regarding my sensitive data.	Kehr, Kowatsch, Wentzel & Fleisch (2015)
	TRSD_2	This data marketplace would tell the truth and fulfill promises related to my sensitive data.	
	TRSD_3	I expect this data marketplace would be honest with me regarding the sensitive data I would provide.	
Trust in data buyers	TRSB_1	I expect that data buyers would be trustworthy in handling the data they got from this data marketplace.	Kehr, Kowatsch, Wentzel & Fleisch (2015)
	TRSB_2	I expect that data buyers would tell the truth and fulfill promises in handling the data they got from this data marketplace.	
	TRSB_3	I expect that data buyers would be honest when handling the data they got from this data marketplace.	
Willingness to Share data via data marketplaces	WTSD_1	Given the chance, I would share my data via this data marketplace.	Pavlou (2003)
	WTSD_2	Given the chance, I predict that I should share my data via this data marketplace in the future.	
	WTSD_3	It is likely that I will share my data via this data marketplace in the near future.	

Table 34 Survey questions

3.1.3 Sampling

As mentioned, this pre-study is a work in progress and part of the more extensive research. Our population comprises consumers that have a driving license. For this pre-study, we used a custom pre-screening in Prolific to restrict our samples to individuals from 18 years old and older as this is the minimum age to have a driving license in most countries. We also exclude participants from the United Kingdom (UK nationalities or those currently living in the UK) as this will be our sample for the main study. Furthermore, we also exclude participants who already took part in our other studies (e.g., D2.6, study 1, and study 3 of this deliverable) to ensure the reliability of the answers of our participants.

Table 35 presents the demographic characteristics of the conducted sample. We conducted the data collection on 9 September 2021, and we managed to recruit a sample of 300 participants (165 male, 126 female, nine others/prefer not to say). The average age of participants was 30.1 years old (SD = 8.87),

and about 70.7% of them are part of the younger generation (18-34 years old). Most of them reside in the United States (53.3%), France (20.7%), and South Africa (6.7%). The majority had already finished a graduate degree (35%), followed by an undergraduate degree (30.3%) and high school diploma/A-level education (18%). More than half of the participants currently work full-time (56%) or part-time (13.7%) and primarily work in the IT (19.7%) or finance industry (7.3%). About one-third of our participants hold a managerial position, either at a junior (5.7%), middle (18%), or upper management level (9.3%). In terms of access to and ownership of cars, only 10% of participants did not have access at all. The rest are either own a car (63.7%), have access via family members (22.3%), or have access via leasing or rental (4%). Additionally, we also asked participants about their familiarity with data marketplaces and privacy-preserving technologies. Interestingly, 53.4% of participants claimed that they are familiar with data marketplaces. However, only 23% of participants have prior knowledge about privacy-preserving technologies before taking part in the survey. Further, we found that the majority of our participants are privacy pragmatists (60.3%), followed by privacy unconcerned (27.3%) and privacy fundamentalists (12.3%). This division is broadly similar to the distribution of privacy perspectives in the general population.

Variable	Demographic	N	%
Age	18-24	108	36%
	25-34	104	34.6%
	35-44	67	22.3%
	45-54	15	5%
	>54	6	2%
Gender	Male	165	55%
	Female	126	42%
	None of the above	7	2.3%
	Prefer not to say	2	0.6%
Education level	Doctorate degree (Ph.D./other)	15	5%
	Graduate degree (MA/MSc/MPhil/other)	105	35%
	Undergraduate degree (BA/BSc/other)	91	30.3%
	Technical/community college	24	8%
	High school diploma/A-levels	54	18%
	Secondary education (e.g., GED/GCSE)	7	2.3%
	Prefer not to say	3	1%
	I do not know/not applicable	1	0.3%
Car ownership	Yes	91	63.7%
	Have access via leasing/rental	12	4%
	Have access via parents/family	67	22.3%
	No	30	10%
Experience with data marketplaces	Shared data through data marketplaces multiple times	29	9.7%
	Shared data through data marketplaces once	27	9%
	Know data marketplaces but never shared data through it	104	34.7%
	Never heard of data marketplaces	140	46.7%
Familiarity with privacy-preserving technologies	Already know before the survey	69	23%
	Have some idea because of the survey	184	61.3%
	Still have no idea after the survey	47	15.7%
Westin Privacy Segmentation Index	Privacy fundamentalists	37	12.3%
	Privacy unconcerned	82	27.3%
	Privacy pragmatists	181	60.3%

Table 35 demographic characteristics (N=300)

3.2 Results

3.2.1 Confirmatory Factor Analysis

We conducted a Confirmatory Factor Analysis (CFA) to validate our constructs and measurement model (Brown & Moore, 2012). Through five rounds of analysis, we assessed the model fit, construct validity, and identified areas of misfit (modification indices). To assess the model fit, we used measures like the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA). We followed a suggestion by Hu and Bentler (1999) for a good fit of those three measures: both CFI and TLI should be 0.95 or higher, and RMSEA should be 0.6 or lower. The results show a good level of the fit index of the model, with CFI = 0.988, TLI = 0.984, and RMSEA = 0.043.

Next, we looked into standardized estimates of each survey item using a threshold of 0.70 (Fornell & Larcker, 1981) and removed two items that did not meet the criteria (CTRL_4 and RISK_2). Then, we assessed the internal reliability of our model by looking at the Composite Reliability (CR) of each construct, which should have a value of 0.7 or higher. Subsequently, we assessed convergent validity through the Average Variance Extracted (AVE), which should be greater than 0.5 (Fornell & Larcker, 1981). As seen in Table 36, we established internal reliability and convergent validity.

Factor	Indicator	Standardized estimates	Mean	Std. dev	R^2	AVE	CR
Perceived Control	CTRL_1	0.845	3.137	1.240	0.715	0.653	0.849
	CTRL_2	0.824	3.250	1.202	0.679		
	CTRL_3	0.752	2.707	1.322	0.566		
Privacy Concerns	PRIV_2	0.816	3.233	1.221	0.667	0.723	0.839
	PRIV_3	0.883	3.380	1.206	0.780		
Perceived Risk	RISK_1	0.855	3.100	1.172	0.732	0.761	0.864
	RISK_3	0.880	3.107	1.186	0.774		
Trust in data marketplaces operator	TRSD_1	0.830	3.763	0.926	0.690	0.724	0.887
	TRSD_2	0.861	3.633	0.991	0.742		
	TRSD_3	0.862	3.713	0.991	0.743		
Trust in data buyers	TRSB_1	0.925	3.443	1.133	0.855	0.869	0.952
	TRSB_2	0.946	3.423	1.090	0.895		
	TRSB_3	0.926	3.477	1.134	0.858		
Willingness to Share data	WTSD_1	0.923	3.410	1.083	0.851	0.846	0.942
	WTSD_2	0.926	3.300	1.126	0.858		
	WTSD_3	0.909	3.287	1.190	0.827		

Table 36 Descriptive statistics, convergent validity, internal consistency, and reliability

We also examined the discriminant validity of the constructs. To do this, we checked whether the correlation among constructs is lower than the square root of AVE (Fornell & Larcker, 1981). As Table 37 shows, all inter-construct correlation coefficients are well below the square root of AVE, meaning that discriminant validity was established.

Factors	Perceived Control	Privacy Concerns	Perceived Risk	Trust in data marketplaces operator	Trust in data buyers	Willingness to Share data
Perceived Control	**0.808**					
Privacy Concerns	-0.260	**0.850**				
Perceived Risk	-0.222	0.758	**0.872**			
Trust in data marketplaces operator	0.514	-0.423	-0.452	**0.851**		
Trust in data buyers	0.524	-0.379	-0.396	0.793	**0.932**	
Willingness to Share data	0.466	-0.406	-0.502	0.676	0.691	**0.920**

Table 37 Discriminant validity: correlation among constructs and the square root of AVE

Furthermore, we assessed modification indices to identify cross-loadings, which are items that load on other constructs due to high correlations between the items from two different constructs. After five rounds of analysis, we removed two items (PRIV_1 and PRIV_4) as both have very high modification indices (higher than 10). We also removed RISK_4 because it was cross-loaded in all other constructs than **Perceived Risk**. Our final model comprised of five factors and 16 items (see Table 36).

In the last step, we conducted Multi-Group Confirmatory Factor Analysis (MGCFA) to see if we could compare different groups (see section 3.1.1). We estimated the model using configurable invariance testing and found a good level of the fit index, with CFI = 0.980, TLI = 0.975, and RMSEA = 0.053. All groups also show convergent validity and discriminant validity, with all standardized estimates higher than 0.7, CR higher than 0.7, and AVE higher than 0.5. Moreover, the comparison between the square root of AVE and all inter-construct correlation coefficients in all groups suggests discriminant validity. In addition, modification indices in all groups were not an issue as we found no very high modification indices and no items that were cross-loaded in all other constructs.

3.2.2 Comparing the effect of three data sharing scenarios

Before we proceed with further analysis, we need to determine the score of each construct. To do so, we aggregate the score of items that belong to each construct and divide those values by the number of items (see Table 36). For instance, **Perceived Control** consisted of three items (CTRL_1, CTRL_2, and CTRL_3). Hence, we computed a new variable in the dataset (CTRL_Avg) by calculating the average of these three items. The same approach also applies to other factors, and we will use these new variables for the remainder of the analysis.

Factors	Group	N	Mean	Median	Std. Dev
Perceived Control	TTP	100	2.78	2.5	1.143
	MPC	100	3.16	3.333	1.009
	DCP	100	3.153	3.333	1.104
Privacy Concerns	TTP	100	3.615	4	0.953
	MPC	100	3.145	3	1.153
	DCP	100	3.16	3	1.202
Perceived Risk	TTP	100	3.35	3.5	0.991
	MPC	100	2.95	3	1.134
	DCP	100	3.01	3	1.148
Trust in data marketplaces operator	TTP	100	3.513	3.67	0.918
	MPC	100	3.796	4	0.858
	DCP	100	3.8	4	0.826
Trust in data buyers	TTP	100	3.163	3	1.067
	MPC	100	3.61	4	1.044
	DCP	100	3.57	4	1.048
Willingness to Share data	TTP	100	3.043	3.33	1.106
	MPC	100	3.57	4	0.937
	DCP	100	3.383	4	1.109

Table 38 Descriptive statistics for all factors across three data sharing scenarios

Table 38 present the descriptive statistics of all factors across three different data-sharing scenarios. The higher mean score indicates better performance, except for the privacy risk and **Perceived Risk** constructs. We found that, on average, MPC and DCP scores are higher in all aspects compared to TTP, but MPC and DCP seem to have a similar average score in most factors. MPC scores slightly higher than DCP in **Perceived Control**, trust in data buyers, and **Willingness to Share data**. Meanwhile, DCP scores slightly higher than MPC in Privacy Concerns, **Perceived Risk**, and trust in data marketplaces operator, although the differences seem very small. Nevertheless, further statistical analyses are needed to confirm that the differences between groups are indeed significant.

In the next step of the analysis, we wanted to compare the effect of three data sharing scenarios (TTP, MPC, and DCP, see section 3.1.1) on all factors (see Table 39). To determine which tests are appropriate, we first performed Levene's test to see if variances are equal in all conditions for all factors. In two of the factors, namely Privacy Concerns ($p = 0.009$) and **Willingness to Share data** ($p = 0.018$), variances in all groups are not equal, meaning that we can only use a non-parametric test (Kruskal-Wallis test). The rest of the factors met the criteria of equal variances in all groups, making it appropriate to use a one-way ANOVA for the analysis.

Based on a one-way between-subjects ANOVA, we found a significant effect of different data sharing scenarios on **Perceived Control** [$F(2,297) = 4.00$, $p = 0.019$, $\omega^2 = 0.02$], **Perceived Risk** [$F(2,297) = 3.89$, $p = 0.021$, $\omega^2 = 0.019$], trust in data marketplaces operator [$F(2,297) = 3.59$, $p = 0.029$, $\omega^2 = 0.017$], and trust in data buyers [$F(2,297) = 5.51$, $p = 0.004$, $\omega^2 = 0.029$] at the $p < .05$ level for the three scenarios. Post hoc testing using Tukey's correction were used to identify where the group differences are. We found that **Perceived Control** in both MPC ($p = 0.037$) and DCP scenarios ($p = 0.042$) are greater than in the TTP scenario. However, we found no significant differences between MPC and DCP scenarios ($p = 0.999$). Similarly, participants in both MPC ($p = 0.008$) and DCP ($p = 0.018$) scenarios perceived a higher degree of Trust in data buyers than those in the TTP scenario. However, we also found no significant differences between MPC and TTP scenarios ($p = 0.961$). Meanwhile, we found no significant differences in terms of **Perceived Risk** in data sharing between DCP and TTP scenarios (p

= 0.073) as well as between DCP and MPC scenarios (p = 0.920). But, participants in the TTP scenario perceived a higher degree of risk in data sharing than those in the MPC scenario (p = 0.027). Finally, concerning Trust in data marketplaces operator, we found no significant differences between participants in all three scenarios.

Factors	One-way ANOVA	Post Hoc Test		
		Group comparison	Mean Difference	p_{tukey}
Perceived Control	F(2,297) = 4.004, p = 0.019*	TTP - MPC	-0.38	0.037*
		TTP - DCP	-0.373	0.042*
		MPC - DCP	0.007	0.999
Perceived Risk	F(2,297) = 3.893, p = 0.021*	TTP - MPC	0.4	0.027*
		TTP - DCP	0.34	0.073
		MPC - DCP	-0.06	0.92
Trust in data marketplace operators	F(2,297) = 3.588, p = 0.029*	TTP - MPC	-0.283	0.057
		TTP - DCP	-0.287	0.053
		MPC - DCP	-0.004	0.999
Trust in data buyers	F(2,297) = 5.514, p = 0.004*	TTP - MPC	-0.447	0.008**
		TTP - DCP	-0.407	0.018*
		MPC - DCP	0.04	0.961

Table 39 Results of one-way ANOVA with post hoc group comparison

Furthermore, a Kruskal-Wallis test reveal that Privacy Concerns [H(2) = 10.102, p = 0.006] and **Willingness to Share data** [H(2) = 12.030, p = 0.002] were significantly affected by different data sharing scenarios (see Table 40). Pairwise comparisons showed that participants in both MPC (p_{holm} = 0.006) and DCP scenarios (p_{holm} = 0.008) perceive lower **Privacy Concerns** compared to the TTP scenario, but we found no significant differences between MPC and DCP scenarios (p_{holm} = 0.411). Both MPC and DCP also significantly increase participants' **Willingness to Share data** (p_{holm} = 0.001 and p_{holm} = 0.022 respectively) compared to the TTP. However, there were no significant differences between MPC and DCP scenarios (p_{holm} = 0.133).

Factors	Kruskall-Wallis Test	Dunn's Post Hoc Test		
		Group	Median	p_{holm}
Privacy Concerns	H(2) = 10.102, p = 0.006*	TTP	4	0.006**
		MPC	3	0.008**
		DCP	3	0.411
Willingness to Share data	H(2) = 12.030, p = 0.002*	TTP	3.33	0.001**
		MPC	4	0.022*
		DCP	4	0.133

Table 40 Kruskall-Wallis Test

3.3 Conclusions

In this study, we have investigated the effect of MPC on individuals' Willingness to Share data through data marketplaces within a specific context of connected car data. First, using confirmatory factor analysis, we found that Perceived Control, Privacy Concerns, Perceived Risk, trust in data marketplaces operator, and trust in data buyers are important indicators to measure Willingness to Share data in privacy-preserving data marketplaces. The model also provides a good fit based on our data.

Second, we found that, when introduced to a new privacy-preserving approach in data sharing (i.e., MPC), people would be more willing to share their data compared to a conventional solution (i.e., TTP). This is because MPC, and even made-up technology like DCP, could increase individuals' feeling of control over data and trust towards data buyers. Moreover, individuals could perceive lower risk and concerns regarding privacy while sharing data, ultimately increasing their Willingness to Share. However, we found no differences between MPC and DCP in all factors, suggesting that citizens might not care in more detail about the technology being used. Furthermore, we found no differences concerning trust in data marketplaces operators. This would mean that technical solutions like MPC might not be relevant in increasing the trustworthiness of data marketplaces. Nevertheless, individuals are still willing to share their data.

4. Study 3: Relative importance of MPC architectures in data sharing

The second study provided further insights on the potential of Safe-DEED technologies in becoming an alternative to existing approaches to data sharing. Still, MPC can be deployed in various configurations, and each of them comes with different trade-offs for users, as we have seen in the deliverables of T2.2. Hence, we first need to assess the relative importance of MPC deployment scenarios for users. Then, we need to identify the most optimal MPC configuration that might result in the highest users' Willingness to Share data.

This section summarizes findings from an MSc thesis that was part of WP2[3], in which we conducted a stated choice experiment with individual citizens in the context of sharing driving data through MPC-enabled data marketplaces[4]. Specifically, this study aimed to investigate car users' preferences and trade-offs between different aspects before sharing their driving data. We focused on assessing the relative importance of MPC architecture to three other important and measurable factors: risk of data disclosure, social influence, and monetary benefits (see Table 41).

Factors	Description	Selected relevant studies
Risk of data disclosure	The probability of the number of incidents per 100 that were hypothetically probable when sharing driving data on MPC-enabled data marketplaces.	Koch et al. (2021); Krasnova et al. (2010); Skatova et al. (2013); Trabelsi et al. (2009); Xie et al. (2006)
MPC architecture	How MPC is deployed in data marketplaces, either centralized (MPC is installed in a central server, the data is stored and processed in a central server) or decentralized form (MPC is installed in the users' car, the data is processed in the car).	Agahari et al. (2021); Archer et al. (2018); Dhillon (2015)
Social influence	The estimation of other consumers that they know also partake in data sharing through MPC-enabled data marketplaces.	Chiregi & Navimipour (2016); Mattke et al. (2020); Sun (2013)
Benefit	Financial incentives that consumers receive every month for sharing their driving data on MPC-enabled data marketplaces.	Hann et al. (2007); Jen et al. (2013); Derikx et al. (2016)

Table 41 Definition of factors included in study 3

While there are similarities between study 2 and study 3 in terms of the context (i.e., sharing driving data through data marketplaces), both studies are essentially different (see Table 1). In study 2, we compared different technical solutions (TTP, MPC, and DCP) and investigated whether MPC performs better than the others. In study 3, we only focused on MPC-based data marketplaces and investigated which MPC configurations and other conditions are the most preferable for car users before sharing their driving data through data marketplaces. This is important because (1) prior research in Safe-DEED (Agahari, Dolci & de Reuver, 2021) found that the different ways to implement MPC in data

3 The thesis is accessible at: http://resolver.tudelft.nl/uuid:9d4303cb-be3d-4964-b89e-c9736d840fcc

4 The data underlying this study is available at: http://resolver.tudelft.nl/uuid%3A9d4303cb-be3d-4964-b89e-c9736d840fcc

marketplaces affect the business model, and (2) the economic modeling in D2.5 considers multiple configuration scenarios for MPC as well. This makes study 3 relevant to explore which configuration would result in the highest Willingness to Share data.

4.1 Method

4.1.1 Choice experiment

We conducted a Stated Choice Experiment (SCE) to let people choose between MPC-enabled data marketplace options instead of asking the willingness directly. This approach is advantageous due to its flexibility: we are not bound by existing data marketplaces because of the experimental design (Mandeville, Lagarde, and Hanson, 2014). Furthermore, Discrete Choice Modeling (DCM) is a mathematically elegant way to determine preferences for certain MPC-enabled data marketplaces because it shows people's preferences of attributes without asking them these difficult questions directly. Apart from these advantages, one must be aware of the hypothetical bias in this type of modelling and the respondents (Rakotonarivo, Schaafsma, and Hockley, 2016). Due to the experimental nature of this approach, people's choices might not reflect what people would choose in reality. One should also be aware of the constant trade-off between the reliability and validity of DCM reflected in the results (Carson et al., 1994).

The experiment consisted of four parts. After briefly introducing the survey's purpose and asking for consent, we presented a short introductory video explaining how MPC works and an illustrative example. Next, we introduced participants to the scenario where they are asked to exchange personal car data via MPC-enabled data marketplaces to help navigation companies improve their services (i.e., road suggestions). Subsequently, in the Stated Choice Experiment, we asked participants to choose between three alternatives each time while considering the previously introduced scenario. Each participant received the same nine options, combining the different levels of four factors (see Table 42). Afterward, we asked demographic questions like gender, age, nationality, and employment. We also asked participants' familiarity with privacy-preserving technologies and data marketplaces. We conclude the survey by asking three privacy-related questions to determine the Westin's Privacy Segmentation Index of each participant (see Section 3.1.2).

Factors	Levels
Risk of data disclosure	1. Low: the consumers are exposed to 1 incident on 100 occasions. 2. Moderate: the consumers are exposed to 5 incidents on 100 occasions. 3. High: the consumers are exposed to 10 incidents on 100 occasions.
MPC Architecture	1. The MPC protocol is installed centrally at data marketplaces. Your car data is transferred to the central MPC computation server hosted by the data marketplace operator. The computation is performed centrally. 2. The MPC protocol is installed in your car. Your car data stays with you. The computation is performed in your car.
Social Influence	1. Hardly anyone you know uses this technology when sharing data on data marketplaces. 2. About half of the people you know use this technology when sharing data on data marketplaces. 3. Almost all the people you know use this technology when sharing data on data marketplaces.
Benefit	1. Participants receive no benefit for inputting their driving data. 2. Participants receive 10 dollars per month for inputting their driving data. 3. Participants receive 20 dollars per month for inputting their driving data.

Table 42 Factors and levels included in study 3

We draw upon previous literature to decide upon the levels of each factor. For **Risk of data disclosure,** levels (in terms of percentages) are based on Travisi and Nijkamp (2008), who operationalized environmental accompanied with health risk levels in the Italian agriculture sector. Hauber et al. (2013) did materialize risk in the same way but in the field of healthcare, specifically risk in medicines. As MPC is not yet in operation, no relevant sources state the amount of data disclosures. Therefore, we follow Koch, Krenn, Pellegrino, and Ramacher (2021), who initialized a table that gives an overview of the threats in data sharing. Own assumptions are made about numerating low, medium, or high likelihoods of a threat. Data disclosure due to sharing car data by MPC on data marketplaces is assumed to be low if it happens in 1% of the times people share. Medium is assumed to be 5% and high 10% of the time.

For **MPC Architecture**, two levels are operationalized. Following Archer et al. (2018), MPC is decentralized by design, which means that this relates to a decentralized architecture where the driving data stays at the car during the computation process. Nevertheless, we differentiate between a centralized and decentralized architecture. Participants need to choose whether the data is stored and processed on a central server or in their car to allow us to retrieve the consumers' preferences regarding the architecture of MPC. Furthermore, to help participants, we included images to show the difference between centralized and decentralized data control to participants. These levels are consistent with the earlier business model architectures as developed in T2.1.

For the factor **Social Influence**, it is assumed that people follow each other when the consequences are preferable and safe. Such behavior is regulated by feelings and socio-psychological factors that determine sensitivity to social influence (Baddeley, 2010). As there is no research done on stated choice experiments within the context of the social influence in data sharing, we developed our own levels for this factor. We distinguished based on the number of known people that took part in data sharing: hardly anyone, about half the known people, and almost all the known people.

Finally, the factor **Benefit** is a common factor within many choice experiments and lends itself to measuring people's Willingness to Pay (WTP) for certain increases on other factors. The other way around, however, we can show to which degree people are willing to accept a devaluation of a factor in exchange for a (monetary) compensation. In Derikx, De Reuver, and Kroesen (2016), people were willing to compensate their Privacy Concerns on automotive data for an average of 9.54€ per month to insurance companies. This is in the middle of our range between zero and twenty dollars, which is preferable as it adds reliability in estimating this parameter.

The factors and levels are combined in order to generate profiles. Participants then choose between three profiles in each choice task. A full factorial design would yield too many choice tasks to be cognitively manageable for respondents. Hence, a mathematically efficient design is applied to reduce the number of choices a respondent has to make to 9 choice tasks. We used Ngene software to generate the efficient design where standard errors and the required number of choice tasks are minimized. An example of a choice task is provided in Figure 7.

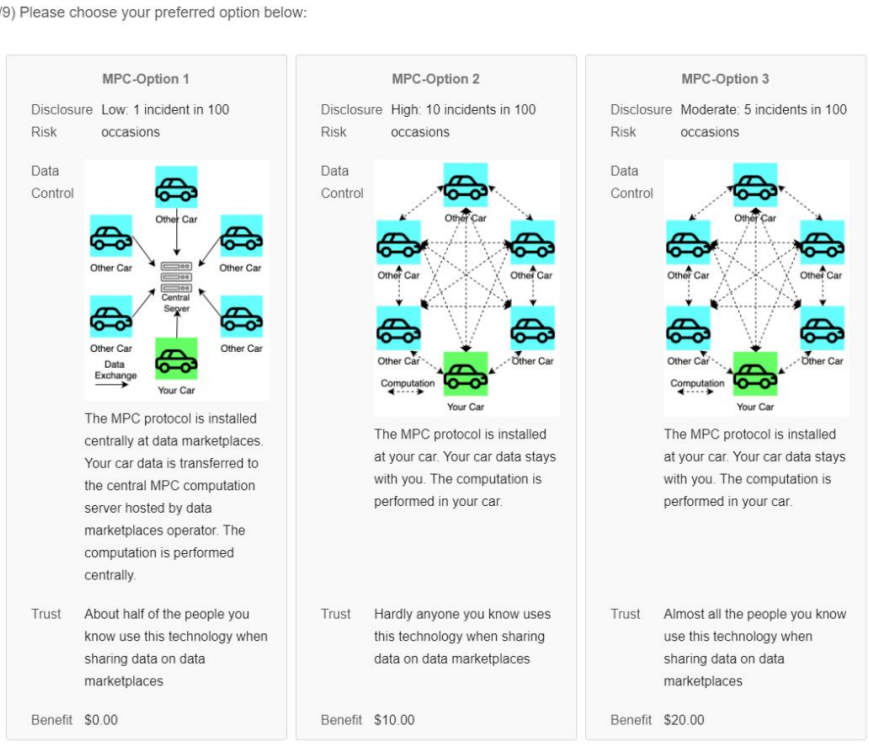


Figure 7 Example of the choice task

4.1.2 Sampling

The targeted group of respondents in this study consists of adults (18+) whose (driving) data will be hypothetically processed. As the targeted population is large and contains all kinds of people, incentives (voluntary) response sampling will be used to obtain sufficient participants. This voluntary response sampling method could lead to bias, as respondents who know more about a subject are more likely to participate in the experiment (Nield and Nordstrom, 2016). However, rewarding the volunteers is aimed to gather an appropriate representation of the population to increase the external validity.

We conducted the data collection on 18 June 2021, and we managed to recruit a sample of 428 participants (216 male, 210 female, two others). About 83% of them are part of the younger generation (18-34 years old). Most of them reside in the United Kingdom (20.8%), South Africa (20.1%), and Portugal (13.3%). Almost half of the participants currently work full-time (28.5%) or part-time (12.9%). Only 19.6% of participants do not have access to a car at all.

We also asked participants about their familiarity with data marketplaces and privacy-preserving technologies. Interestingly, 64.3% of participants claimed that they have at least heard of data marketplaces even though they never used them to share data. Participants seem to have a strong knowledge of privacy-preserving technologies: 61% of them claimed to know the technology before taking part in the survey. Furthermore, we found that the majority of our participants are privacy pragmatists (77.8%), followed by privacy fundamentalists (15.7%), and privacy unconcerned (6.5%).

4.2 Results

We analyzed participants' answers regarding privacy statements before moving further to the choice experiment results (see Table 43). We conducted multiple One-Way ANOVA to determine whether there were any statistically significant differences between the means of three or more independent (unrelated) groups on the privacy statements. We found that older people agree more on privacy

statement 1 (p=0.048), while highly educated people disagree more on privacy statement 2 (p=0.049). We also found that people familiar with data marketplaces agree more on privacy statements 2 (p=0.014) and 3 (p=0.032). Furthermore, we found no significant differences between genders, industry sectors, people with different PPT knowledge, or between different types of car ownership.

Statement	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Mean	Standard deviation
Consumers have lost all control over how personal information is collected and used by companies.	1.6%	12.4%	14%	52.3%	19.6%	3.76	0.961
Most businesses handle the personal information they collect about consumers in a proper and confidential way.	7.7%	34.6%	31.1%	20.6%	6.1%	2.83	1.04
Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.	4.9%	18.9%	36.4%	29.7%	10%	3.21	1.02

Table 43 Privacy statements

Moving to the choice experiments, we presented a summary of the given answers of the processed choice experiment (Table 44) and participants' preferences in a colorized manner (Figure 8). Here, green is preferable, and red is not preferable. By quick inspection, we can see that alternative 2 in the choice set 9 is very dominant and chosen by 84% of the respondents due to its good attribute levels. Unfortunately, this choice set contributes little information due to its dominance.

Choice	Alternative	Alternative 2	Alternative 3
1	19%	14%	**67%**
2	**57%**	23%	20%
3	19%	**47%**	34%
4	25%	**70%**	5%
5	11%	13%	**76%**
6	**67%**	9%	24%
7	29%	18%	**53%**
8	35%	14%	**51%**
9	11%	**84%**	5%

Table 44 Distribution of given DCE choices (N=428)

	Option 1				Option 2				Option 3			
Question	Risk	MPC architecture	Social influence	Benefit	Risk	MPC architecture	Social influence	Benefit	Risk	MPC architecture	Social influence	Benefit
1	Low	Centralized	Moderate	0	High	Decentralized	Low	10	Moderate	Decentralized	High	20
2	Moderate	Centralized	High	20	Low	Centralized	Moderate	0	High	Decentralized	Low	10
3	High	Centralized	High	10	Moderate	Decentralized	Low	20	Low	Decentralized	Moderate	0
4	Moderate	Centralized	Low	20	Low	Decentralized	High	10	High	Decentralized	Moderate	0
5	High	Decentralized	Moderate	0	Moderate	Centralized	Low	10	Low	Centralized	High	20
6	Low	Decentralized	Moderate	10	High	Centralized	High	0	Moderate	Centralized	Low	20
7	High	Decentralized	High	20	Moderate	Centralized	Moderate	0	Low	Centralized	Low	10
8	Low	Decentralized	Low	0	High	Centralized	Moderate	20	Moderate	Centralized	High	10
9	Moderate	Centralized	Low	10	Low	Decentralized	High	20	High	Centralized	Moderate	0

Figure 8 colorized design

Here, we presented the results of the estimated model using the Mixed-Logit Model (ML), where attributes are based on distributions (see Table 45). We found that all factors are highly significant as all p-values < 0.05 and all t-ratios are > 1.96. This model exceeds the null hypothesis with a significance of LRS = 2639.43 and p = 0.000. The rho-squared value is 0.3119, which falls within a range of a good fit between 0.2 and 0.4 (McFadden et al., 1973). Based on the AIC and the BIC values, there exists heterogeneity in attribute taste among people because this estimated model is significantly stronger than the null model.

The ranking of the estimates is calculated through the utility difference of the lowest and highest attribute values. This means that the **Risk of data disclosure** is the most important factor, followed by **Benefit**, **Social Influence**, and **MPC architecture**. Furthermore, the optimal package, which is the highest preferred bundle across respondents and maximizes their preference and utility, is a combination of low risk, MPC installed in own car, almost everybody that participants know uses the technology, and benefits of 10$ per month.

	Beta Estimate	Standard Error	tvalue	pvalue
β_{RISK}	1.26489	0.069933	18.087	0.000
β_{MPC}	0.44651	0.090536	4.932	6.722e07
B_{SOCIAL}	0.47216	0.050470	9.355	0.000
$\beta_{BENEFIT}$	0.09124	0.006165	14.800	0.000
σ_{RISK}	0.96936	0.064806	14.958	0.000
σ_{MPC}	1.51249	0.100607	15.034	0.000
σ_{SOCIAL}	0.76970	0.051689	14.891	0.000
$\sigma_{BENEFIT}$	0.09459	0.006265	15.099	0.000
Number of observations	3852			
0-Loglik	4231.855			
Final-Loglik	2912.14			
LRS /w 0-Loglik	2639.43			
LRS /w MNL	879.21			
Mc Fadden's rho-squared	0.3119			
AIC	5840.28			
BIC	5890.33			

Table 45 ML model estimates

Figure 9 shows the relative importance of the attributes as percentages. In other words, the average measurement of influence an attribute had when the respondents were choosing their preferred alternative. The higher the score, the more weight it carried in the decision-making process (the scores add up to 100%). Each attribute will be discussed, beginning from most important to least important.

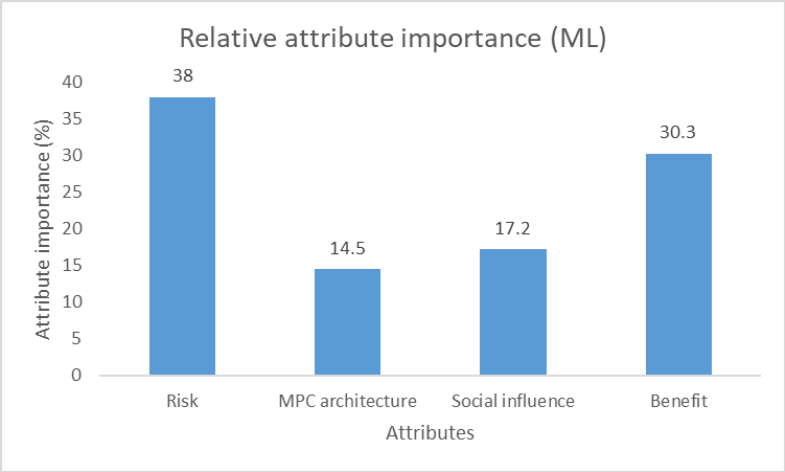


Figure 9 Relative Attribute Importance

The **Risks** estimate (1.2649) a consumer is exposed to by sharing data on data marketplaces will influence the decision to share on a data marketplace the most. This attribute is the most crucial based on the maximum likelihood principle on the filled-in choice data. Relatively, the risk comprises 38.0% of the total importance. Too high risks of data disclosure will eventually lead to discontinuation of sharing (see Figure 10).

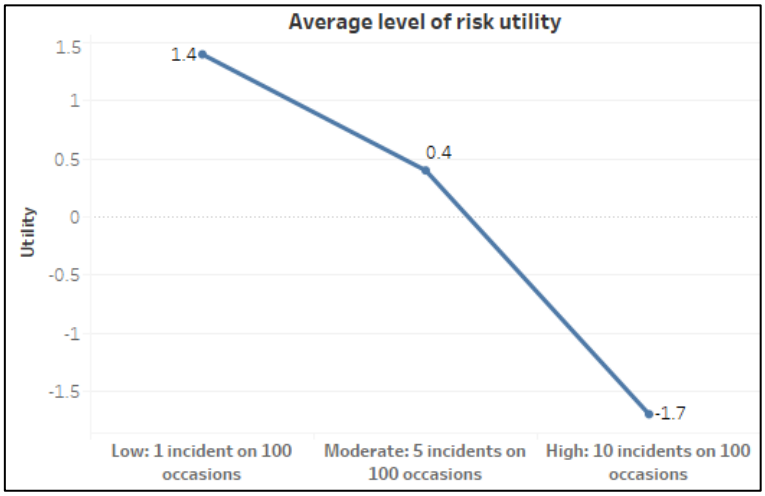


Figure 10 Risk utility

The **Benefits** (0.0969) that the consumer perceives to sharing driving data influence the decision to participate as second most important. Almost a third (30.3%) of choice for a specific platform is based on the amount of benefit people receive. However, as shown in Figure 11, after reaching 10 dollars of benefit per month, the curve flattens, and people experience not as much additional utility for an additional monthly benefit. As this attribute is perceived as very important to respondents, it is crucial to incorporate a certain benefit factor in MPC-enabled data marketplaces.

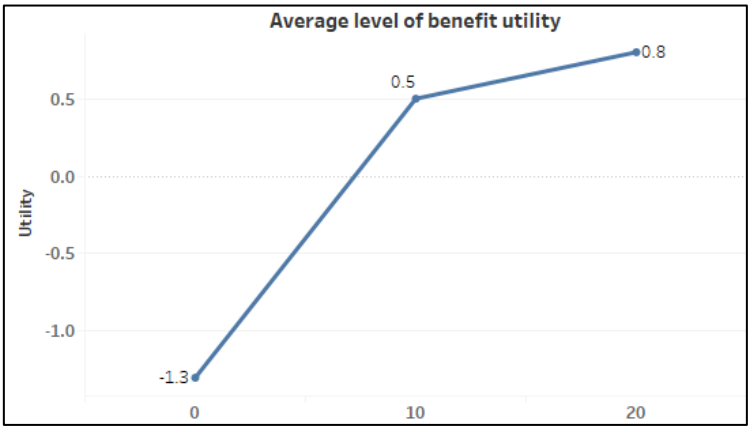


Figure 11 Benefit utility

Social Influence (0.4722) is weighted third most important. About 17.2% of the choices are based on the number of people that participants know to share data through data marketplaces. This has causality with the certain mass of demand a platform needs in order to attract more people. This experiment shows that it is still relatively important that people attract other people to data marketplaces. Moreover, the effect is stronger whenever more people share their driving data using MPCenabled data marketplaces as the average level utility is not fully linear (see Figure 12).

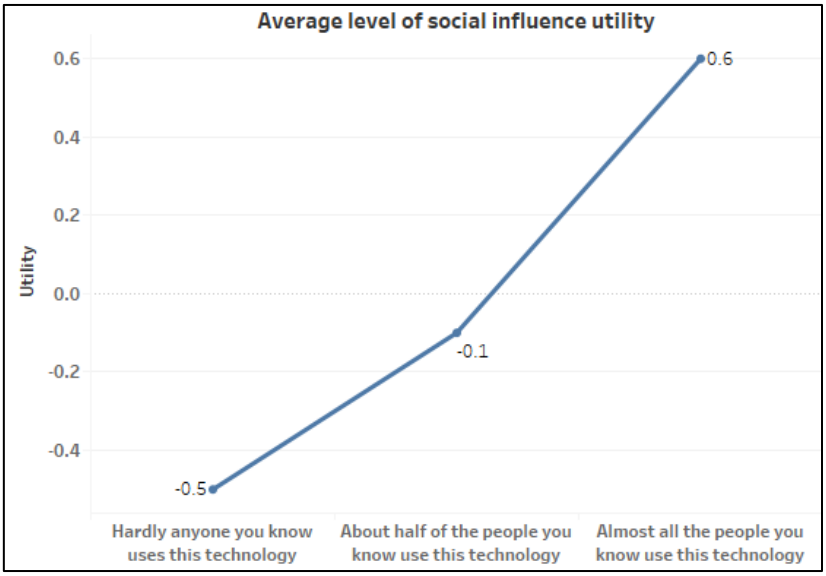


Figure 12 Social influence utility

At last, **MPC architecture** (0.4465) is perceived to be the least important in the choices made. Around 14.5% of choice is based on whether MPC is installed centrally in data marketplaces or installed in the car. It was not expected to be the least important factor, even less important than the herding behavior. People prefer to have a decentralized **MPC architecture**, but it is not that important relative to the other factors as risk or benefit. Furthermore, as this variable only has two levels, linearity cannot be tested (see Figure 13).

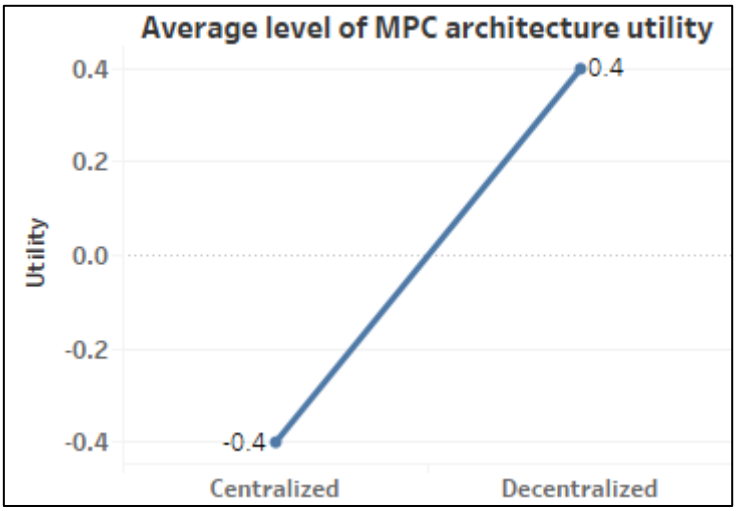


Figure 13 MPC architecture utility

4.3 Conclusions

We found that consumers make trade-offs between various factors when deciding to share data through MPC-enabled data marketplaces. While our choice experiment model did explain a decent amount of choices, our model did not get close to 1, meaning that unexplained variance still exists. This can be explained by additional or different factors that are not concrete and were hard to include in the choice experiment. Either due to complexity or due to being too ambiguous.

Nevertheless, for the included factors in all models, all factors appeared to be relevant in decision making between MPC-enabled data marketplaces. As expected, the risk of data disclosure affected the utility negatively, and all other factors positively affected the utility. We found the risk of data disclosure to be the most important factor among participants, followed narrowly by benefit. Social influence was the third most important factor, and MPC architecture was the least important factor. Furthermore, we showed that taste heterogeneity exists between people regarding the four included factors, as shown by the ML model by the estimated factor sigmas.

Interestingly, the demographic factors and Westin's privacy segmentation index had no significant effect on people's choices in this stated choice experiment. We did find significant differences concerning privacy statements between consumers based on their age, educational level, and experience with data marketplaces. Statistically, older participants feel that consumers lost all control over their personal data compared to younger participants. Meanwhile, highly educated people more often disagree that businesses handle consumer data confidentially and properly than less educated people. Furthermore, the more experienced people with data marketplaces agree more that businesses handle personal consumer information properly and that the existing laws and organizational practices provide enough protection for consumers today.

The key conclusions have a variety of policy consequences for how to set up MPC-enabled data marketplaces. As MPC is not yet massively adopted, modifications can be more easily adapted. As shown, the choice for sharing data via MPC-enabled data marketplaces is mostly dependent on the amount of risk of data disclosure involved and the opportunity to have a certain benefit. The concept of MPC is exactly invented in order to increase safety during sharing of sensitive data. However, it seems important to inform people more about the technology. These educational functions should be applied to governmental authorities such as local authorities, the EU, financial institutions, and academic institutions. As shown in the data, most people do not really know what happens with their personal data and do not know where it is stored or even sold. So, raising awareness regarding the types of risks would be the first task to explain what MPC is.

Furthermore, many participants' comments pointed out that they would like, just as in this survey, to choose how their data is stored and distributed. In other words, have a say in the way the data is handled. By giving people a certain incentive to share their data, people constantly make trade-offs between the amount of risk and the amount of benefit they receive. This incentive helps to gain the appropriate mass of demand, which creates publicity for MPC in general for people to start using this method of hidden sharing. In the end, herding behavior was one of the least important factors. This contradicts literature on trust in data sharing in ecommerce (Kim, Ferrin & Rao, 2009) as people use specific products or share valuable data generally based on trust. However, this research gave insights into the way people herd in terms of online automotive data sharing.

5. Discussion and conclusions

The main objective of this deliverable is to provide quantitative evidence on how the developed privacy-preserving technologies within Safe-DEED affect citizen trust and intention to use. Based on the three studies conducted, we found that Safe-DEED technologies could indeed contribute toward trust and intention to use for both businesses and individuals. Such technologies could even create more value beyond trust by inducing the feeling of control over data, reducing the risk of data sharing, and lowering privacy concerns. Hence, our findings provide empirical evidence on the relevance of the Safe-DEED technologies in fulfilling the goals of the data economy, particularly citizen trust, value creation for businesses, adoption of privacy/confidentiality preserving technologies, and ultimately turnover of businesses.

We found the importance of having clear benefits for businesses and individuals before sharing data using Safe-DEED technologies, although in a different form. For business actors, the benefit is ultimately about creating more value and revenue for their business beyond improved privacy and security. This might include new approaches for data analytics and the ability to utilize new data sources. Meanwhile, for individuals, the benefit might simply be about monetary incentives, as shown in study 3. However, it could also mean intangible benefits like improved privacy, security, and having more control over their data. As shown in study 2, people would be more willing to use privacy-preserving approaches (such as MPC protocol developed in Safe-DEED) in data sharing than conventional technology (such as Trusted Third Party). This is because they feel the benefit of using MPC, such as a stronger feeling of control over data, reduced risk, and lower Privacy concerns. This finding is in line with our earlier work in D2.6, in which we found relative advantage as one of the factors influencing Willingness to Share data through MPC-based applications.

We also found the importance sof having a clear explanation and workflow on how Safe-DEED technology works, what kind of results the users will get, and how useful it will be for users. This is necessary as we found constraints regarding the complexity of the technology. The usefulness and benefits of the demonstrator do not necessarily lead to more adoption of the technology, especially if it is difficult to use and understand. This is consistent with our findings in D2.4: if visualized and appropriately explained, the demonstrator could contribute to perceptions of trust, security, and control, ultimately influencing willingness to use it.

At the same time, we have gained interesting insights into people's attitudes towards Safe-DEED technology. It turns out that (potential) users do not really care about the details of the technology and its setup. Prope explanation about the usefulness and the privacy value of the technology would lead to its adoption. Nevertheless, as shown in study 3, people first consider benefits and risks before sharing data. Again, communicating how MPC works plays an important role here. Since the purpose of the technology is precisely about reducing the risk of data sharing and creating benefits in terms of higher control over data, it is crucial to raise awareness of the usefulness of this technology.

Additionally, we found contrasting findings concerning the target group of the Safe-DEED technology. While study 1 suggests that certain groups are more willing to use technology than others, study 3 indicates no differences in preferences for different groups. This implies that if Safe-DEED technologies would be offered to businesses, it is better to target people in the data-related role and product development, especially since they would require more data from other sources that they could not get before. In contrast, if MPC is implemented in consumer-facing platforms, it does not matter who the target users will be, as long as it is explained that it is useful. This goes back to the importance of properly communicating the technology to a wider audience.

Furthermore, our findings in study 2 imply that increasing trust towards Safe-DEED technologies requires approaches beyond technical solutions like MPC. This is because it could only increase trust towards the data requesting party and not trust towards the application. Given the complex nature of trust and its importance in the data economy, we cannot simply rely on technical solutions alone. Proper

data governance mechanisms are still necessary to complement technical means like MPC, covering aspects like what kind of data can be accessed, who can access those data, and the purpose of using the data.

Despite our contribution to this deliverable, several limitations should be taken into account. First, we cannot generalize our findings to all populations, as our samples in all studies are skewed towards younger and more technical people. This is likely because we rely on Prolific as a tool to collect responses, as most of the registered users are young people. Moreover, using Prolific in study 1 might be an issue since we focused on the business perspective, while users in Prolific are individuals and do not represent organizations per se. However, we mitigated this constraint by introducing a persona and ensuring that only those with managerial positions could participate in our study. Further research might benefit from replicating the study with real-life business actors.

The second limitation is about the mock-ups and scenarios introduced in each study. While we use a working prototype in study 1, we only use hypothetical scenarios and mock-ups in both studies 2 and 3. This is mainly due to the context we focus on in both studies, which is about data marketplaces. At present, personal data marketplaces are still limited, let alone MPC-enabled data marketplaces. Therefore, we can only present screenshots and possible scenarios of MPC in data marketplaces and rely on participants' understanding to answer our survey questions. Put differently, we evaluate how we visualize and explain the technology, not the technology itself. Similarly, in study 1, participants likely put more attention on the user interface of the demonstrators rather than the underlying Safe-DEED technologies. Nevertheless, this is likely to be the case with MPC and related technologies like blockchain, which runs in the background and cannot be verified by end-users.

In conclusion, Safe-DEED technologies could contribute toward trust and intention to share data of both businesses and individuals. Such technologies could even create more value beyond trust by inducing the feeling of control over data, reducing the risk of data sharing, and lowering Privacy concerns. Nevertheless, realizing these contributions would require a proper communication strategy to explain how it works and the benefits that users will get by using the technology.

To sum up, we propose three recommendations based on our findings in this deliverable. First, a proper communication strategy should be established to explain how Safe-DEED technologies work and the benefits users will get. This might be in the form of improved user experience (UX) or user interface (UI) of the demonstrator, clearly showing what is going on at the backend or raising awareness towards general audiences. Second, in targeting potential users, the focus should be on the knowledgeable groups, such as those involved in new product development and highly engaged in data-related activities. These groups are likely to be able to 'see through' the technology and appreciate its value, making it crucial to get them on board first before expanding to other potential users. And third, as a technical solution, Safe-DEED technologies could only solve some data sharing barriers, such as Privacy concerns and control over data. Therefore, a comprehensive approach is necessary by complementing those technologies with proper data governance mechanisms.

6. References

Adams, T. L., Li, Y., & Liu, H. (2020). A Replication of Beyond the Turk: Alternative Platforms for Crowdsourcing Behavioral Research – Sometimes Preferable to Student Groups. *AIS Transactions on Replication Research*, *6*, 1–22. https://doi.org/10.17705/1atrr.00058

Agahari, W., Dolci, R., & de Reuver, G. A. (2021). Business model implications of privacy-preserving technologies in data marketplaces: The case of multi-party computation. In *29th European Conference on Information Systems (ECIS 2021): Human Values Crisis in a Digitizing World* (pp. 1-16). Association of the Information Systems (AIS). https://aisel.aisnet.org/ecis2021_rp/59/

Archer, D. W., Bogdanov, D., Lindell, Y., Kamm, L., Nielsen, K., Pagter, J. I., Smart, N. P., & Wright, R. N. (2018). From Keys to Databases—Real-World Applications of Secure Multi-Party Computation. *The Computer Journal*, *61*(12), 1749–1771. https://doi.org/10.1093/comjnl/bxy090

Baddeley, M. (2010). Herding, social influence and economic decisionmaking: Sociopsychological and neuroscientific analyses. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *365*, 281–90. https://doi.org/10.1098/rstb.2009.0169

Brown, T. A., & Moore, M. T. (2012). Confirmatory factor analysis. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 361–379). The Guilford Press.

Carson, R. T., Louviere, J. J., Anderson, D. A., Arabie, P., Bunch, D. S., Hensher, D. A., Johnson, R. M., Kuhfeld, W. F., Steinberg, D., Swait, J., Timmermans, H., & Wiley, J. B. (1994). Experimental analysis of choice. *Marketing Letters*, *5*(4), 351–367. https://doi.org/10.1007/bf00999210

CBS. (2019). Unemployment - *The Netherlands on the European scale*. CBS. Retrieved November 8, 2021, from https://longreads.cbs.nl/european-scale-2019/unemployment/

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of economic behavior & organization*, 81(1), 1-8.

Chiregi, M., & Navimipour, N. J. (2016). A new method for trust and reputation evaluation in the cloud environments using the recommendations of opinion leaders' entities and removing the effect of troll entities. *Computers in Human Behavior*, *60*(C), 280–292. https://doi.org/10.1016/j.chb.2016.02.029

Cranmer, G. (2017). One-Group Pretest–Posttest Design. In M. Allen (Eds.), *The SAGE Encyclopedia of Communication Research Methods* (pp. 1125–1126). https://doi.org/10.4135/9781483381411.n388

Derikx, S., de Reuver, M., & Kroesen, M. (2015). Can Privacy Concerns for insurance of connected cars be compensated? *Electronic Markets*, *26*(1), 73–81. https://doi.org/10.1007/s12525-015-0211-0

Dhillon, G. (2015). What to do before and after a cybersecurity breach. *American University, Washington, DC, Kogod Cybersecurity Governance Center*. https://doi.org/10.17606/yqqs-jr09

Dimitrov, D. M., & Rumrill, P. D. (2003). Pretest-posttest designs and measurement of change. *Work*, *20*. 159–165. https://pubmed.ncbi.nlm.nih.gov/12671209/

Dinev, T., & Hart, P. (2006). An Extended Privacy Calculus Model for E-Commerce Transactions. *Information Systems Research*, *17*(1), 61–80. https://doi.org/10.1287/isre.1060.0080

Fornell, C., & Larcker, D. F. (1981). Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics. *Journal of Marketing Research, 18*, 382-388. http://dx.doi.org/10.2307/3150980

Gordon, R. A. (2015). *Regression Analysis for the Social Sciences* (Second ed.). Routledge.

Hann, I. H., Hui, K. L., Lee, S. Y. T., & Png, I. P. (2007). Overcoming Online Information Privacy Concerns: An Information-Processing Theory Approach. *Journal of Management Information Systems*, *24*(2), 13–42. https://doi.org/10.2753/mis0742-1222240202

Hauber, A., Arden, N., Mohamed, A., Johnson, F., Peloso, P., Watson, D., Mavros, P., Gammaitoni, A., Sen, S., & Taylor, S. (2013). A discrete-choice experiment of United Kingdom patients' willingness to risk adverse events for improved function and pain control in osteoarthritis. *Osteoarthritis and Cartilage*, *21*(2), 289–297. https://doi.org/10.1016/j.joca.2012.11.007

Jen, W., LU, M. L., Wang, W.T., & Chang, Y.T. (2013). Effects of Perceived Benefits and perceived costs on passenger's intention to use selfticketing kiosk of taiwan high speed rail corporation. *Journal of the Eastern Asia Society for Transportation Studies*, *10*, 215–230. https://doi.org/10.11175/easts.10.215

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

Kim, D. J., Ferrin, D. L., & Rao, H. R. (2009). Trust and satisfaction, two stepping stones for successful ecommerce relationships: *A longitudinal exploration. Information systems research*, *20*(2), 237–257. https://doi.org/10.1287/isre.1080.0188

Koch, K., Krenn, S., Pellegrino, D., & Ramacher, S. (2021). Privacypreserving analytics for data markets using mpc. *arXiv preprint arXiv:2103.03739*

Krasnova, H., Spiekermann, S., Koroleva, K., & Hildebrand, T. (2010). Online social networks: Why we disclose. Journal of information technology, 25(2), 109–125.

Krasnova, H., Spiekermann, S., Koroleva, K., & Hildebrand, T. (2010). Online Social Networks: Why We Disclose. *Journal of Information Technology*, *25*(2), 109–125. https://doi.org/10.1057/jit.2010.6

Kumaraguru, P., & Cranor, L.F. (2005). Privacy Indexes: A Survey of Westin's Studies.

Mandeville, K. L., Lagarde, M., & Hanson, K. (2014). The use of discrete choice experiments to inform health workforce policy: a systematic review. *BMC Health Services Research*, *14*(1). https://doi.org/10.1186/1472-6963-14-367

Mattke, J., Maier, C., Reis, L., & Weitzel, T. (2020). Herd behavior in social media: The role of Facebook likes, strength of ties, and expertise. *Information & Management*, *57*(8), 103370. https://doi.org/10.1016/j.im.2020.103370

McFadden, D. et al. (1973). Conditional logit analysis of qualitative choice behavior.

N., K., Schulze, T., & Veit, D. (2011). *More than fun and money. Worker Motivation in Crowdsourcing – A Study on Mechanical Turk*. AMCIS 2011 Proceedings - All Submissions. https://aisel.aisnet.org/amcis2011_submissions/340/

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*(2018), 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Pavlou, P.A. (2003). Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model. *International Journal of Electronic Commerce 7*(3), 101–134. https://www.jstor.org/stable/27751067

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*(2017), 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*(2017), 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Petronia, M. (2020). *The effect of multiparty computation on firms' willingness to contribute protected data*. https://repository.tudelft.nl/islandora/object/uuid%3Ab0de4a4b-f5a3-44b8-baa4-a6416cebe26f?collection=education

Privitera, G.J. & Ahlgrim-Delzell, L. (2018). *Research Methods for Education*. SAGE Publications

Rakotonarivo, O. S., Schaafsma, M., & Hockley, N. (2016). A systematic review of the reliability and validity of discrete choice experiments in valuing non-market environmental goods. *Journal of Environmental Management*, *183*, 98–109. https://doi.org/10.1016/j.jenvman.2016.08.032

Safe-DEED. (2020. November). *User experiment report v2* (D2.6). https://safe-deed.eu/wp-content/uploads/2020/12/Safe-DEED_D2_6.pdf

Skatova, A., Johal, J., Houghton, R., Mortier, R., Bhandari, N., Lodge, T., Wagner, C., Goulding, J., Crowcroft, J., & Madhavapeddy, A. (2013). Perceived Risks of personal data sharing. *Proc. Digital Economy: Open Digital (Nov. 2013)*.

Sun, H. (2013). A Longitudinal Study of Herd Behavior in the Adoption and Continued Use of Technology. *MIS Quarterly*, *37*(4), 1013–1041. https://doi.org/10.25300/misq/2013/37.4.02

Trabelsi, S., Salzgeber, V., Bezzi, M., & Montagnon, G. (2009). Data disclosure risk evaluation. *2009 Fourth International Conference on Risks and Security of Internet and Systems (CRiSIS 2009)*, 35–72. https://doi.org/10.1109/crisis.2009.5411979

Travisi, C. M., & Nijkamp, P. (2008). Valuing environmental and health risk in agriculture: A choice experiment approach to pesticides in italy. Ecological Economics, 67(4), 598–607. https://EconPapers.repec.org/RePEc:eee:ecolec:v:67:y:2008:i:4:p:598-607

Xie, E., Teo, H. H., & Wan, W. (2006). Volunteering personal information on the internet: Effects of reputation, privacy notices, and rewards on online consumer behavior. *Marketing Letters*, *17*(1), 61–74. https://doi.org/10.1007/s11002-006-4147-1

Xu, H., Dinev, T., Smith, J., & Hart, P. (2011). Information Privacy Concerns: Linking Individual Perceptions with Institutional Privacy Assurances. *Journal of the Association for Information Systems*, *12*(12), 798–824. https://doi.org/10.17705/1jais.00281

Appendix A: Scenarios in online surveys

Experiment 1 Data exchange application

Scenario
Suppose you are a business development manager of a telecom operator selling premium pay-TV packages to customers. Your company is currently a market leader and especially successful in selling live sports TV packages, but you still would like to increase revenues and the subscriber base.

Pre-test: Expectations
Now, suppose you want to identify cross-selling opportunities with one of the major banks in your region. Perhaps there is an area where the bank has many high-value customers. So, you can focus your marketing effort to attract them in subscribing to your premium pay-TV package. To do this, you need to exchange your customer database (CRM) with the bank and vice versa. However, you cannot simply share this customer data because it is privacy-sensitive and highly confidential data.

Suppose that there is a data exchange application that can analyze your customer datasets with datasets from other companies. This analysis resulted in the intersection of these two datasets without revealing the input data to each other. See the image below for an example. Specific marketing efforts could be focused on customer 4 and 5, like intensified marketing strategies, special discounts or a loyalty program.

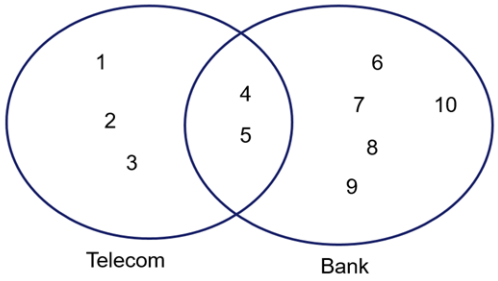
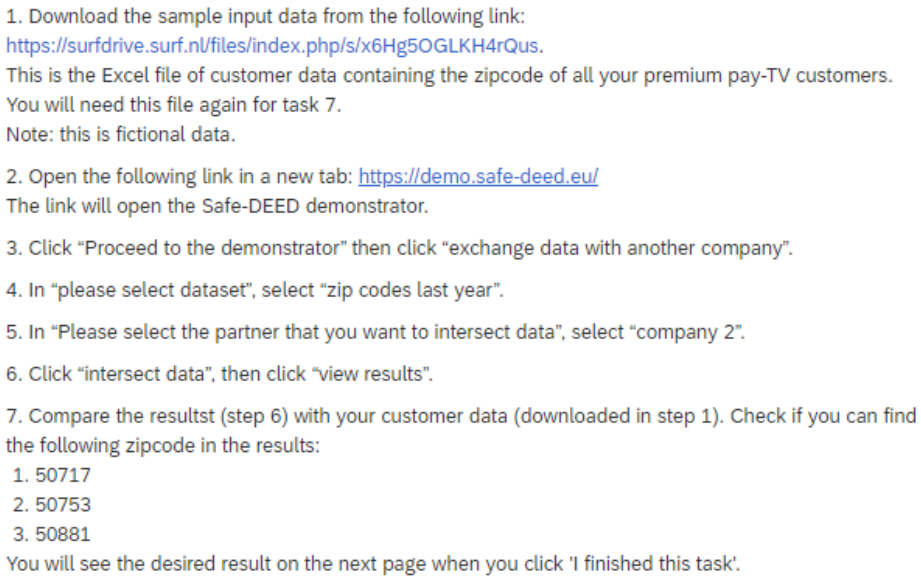


Figure 14 Visualisation of intersection of two datasets

For the following questions, please rate your expectations for such an application.

Treatment: Data exchange application
One application in the Safe-DEED demonstrator, namely the data exchange application, can offer a solution for sharing such confidential information. This application uses a Private Set Intersection (PSI) protocol that computes the intersection of two datasets from different owners without releasing the underlying data. In other words, this application allows you to calculate an intersection of customer data between your company and the bank without giving away your complete customer database (CRM). To better understand how PSI works, please watch a 5-minute introductory video of PSI below (https://youtu.be/S6Bsz2G0stE).

You will now try out the data exchange application in the Safe-DEED demonstrator. Please complete the following tasks and return to this page after you finish each task. After finishing all the tasks, you will be provided a code.


1. Download the sample input data from the following link:
https://surfdrive.surf.nl/files/index.php/s/x6Hg5OGLKH4rQus.
This is the Excel file of customer data containing the zipcode of all your premium pay-TV customers.
You will need this file again for task 7.
Note: this is fictional data.

2. Open the following link in a new tab: https://demo.safe-deed.eu/
The link will open the Safe-DEED demonstrator.

3. Click "Proceed to the demonstrator" then click "exchange data with another company".

4. In "please select dataset", select "zip codes last year".

5. In "Please select the partner that you want to intersect data", select "company 2".

6. Click "intersect data", then click "view results".

7. Compare the resultst (step 6) with your customer data (downloaded in step 1). Check if you can find the following zipcode in the results:
1. 50717
2. 50753
3. 50881
You will see the desired result on the next page when you click 'I finished this task'.

Figure 15 Tasks to experiment with the data exchange applcication

After finishing all the tasks, you will be provided a code.

Post-test: Perceptions of application
In this part, we want to understand your perceptions of the de-anonymization application you just used.

Experiment 2 De-anonymization application

Scenario
In the next pages, you will get questions about an application. To understand this application better, we provide you a scenario. Please read it carefully.
Suppose you are a business development manager of a telecom operator selling premium pay-TV packages to customers. Your company is currently a market leader and especially successful in selling live sports TV packages. But, you still would like to increase revenues and subscriber base.

Post-test: expectation
Now, suppose you want to ask a data analysis consultant to analyze the viewership data. However, since this data is privacy-sensitive, you first need to anonymize it by removing personal information.
Suppose that there is a de-anonymization application that can:

- check if your dataset is sufficiently anonymized.
- anonymize your dataset before sharing to reduce the de-anonymization risks.

For the following questions, please rate your expectations for such an application.

Treatment: De-anonymization application

In the safe-DEED project, we have developed a **de-anonymization application** that can help you to:
1. Check the **de-anonymization risks** of your viewership dataset before sharing it with another party (in this case, a data analysis consultant)
2. Anonymize your viewership dataset through a **k-anonymization** technique

To better understand how the de-anonymization application works, please watch the video below.
https://youtu.be/NSnndxOVV9U

You will now perform a de-anonymization risks analysis of your viewership data. Please complete the following tasks and return to this page after you finish each task.
After finishing all the tasks, you will be provided a code.

1. Download the sample data from the following link: https://surfdrive.surf.nl/files/index.php/s/SwX4yLbnKU7Riow
This is the Excel file of viewership data containing the demographic of all your premium pay-TV customers. You will need this file again for task 9.
Note: this is fictional data.

2. Open the following link in a new tab: https://demo.safe-deed.eu/
The link will open the Safe-DEED demonstrator.

3. Click "Proceed to the demonstrator" then click "provide data for external analysis".

4. Select "de-anonymization risk analysis" then click "company demographic data".

5. Read the explanation of Quasi-Identifiers (QIs), then click "select all" and click "analyze".

6. Read the explanation of results, and move your mouse around the plot until you can see the label. You can read the plot (X, Y) as (number of QIs, probability of de-anonymization in %).

7. Pick one plot and describe (to yourself) what it means using the following formula:
"In my CRM datasets, there are Y% of individuals that are uniquely identifiable by X combinations of Quasi-Identifiers."

8. Click "anonymize" and download the anonymized data from the following link: https://surfdrive.surf.nl/files/index.php/s/jf3BlfmVMaEA0Kr
Note: this is the same data you will get if you click "download anonymized data" in your browser, but the formatting is different.

9. Compare the original data (downloaded in step 1) and the anonymized data (step 8). Observe the differences.

Figure 16 Tasks to experiment with the de-anonymization application

Post-test: perception of application
In this part, we want to understand your perceptions of the de-anonymization application you just used.

Experiment 3 Data valuation application

Scenario
In the next pages, you will get questions about an application. To understand this application better, we provide you a scenario. Please read it carefully.

Suppose you are a business development manager of a telecom operator selling premium pay-TV packages to customers. Your company is currently a market leader and especially successful in selling live sports TV packages. But, you still would like to increase revenues and subscriber base. They have recently started experimenting with adding live streams via Facebook and YouTube just before a match starts but have no insight yet in usage patterns.

Now, suppose you want to ask a data analysis consultant to analyze the usage patterns of live streams data. Assume that the data is already anonymized. Before sharing the data, you want to assess its value and quality.

Pre-test: Expectations
Suppose there is a data valuation application that can help assess the quality, exploitability, and economic value of your data.

For the following questions, please rate your expectations for such an application.

Treatment: Data valuation application
In the safe-DEED project, we have developed a **data valuation application** that can help you assess the quality, exploitability, and economic value of your data. The application works by receiving a structured data set, together with a context and a set of rules for evaluating data quality. The application returns three scores that describe the value of the data:

1. **Qualitative data score:** a score based on the contextual information provided
2. **Automatic data analysis score:** a score based on the quality rules provided
3. **Dataset value:** the aggregate value of the dataset, computed as a mean of the 2 previous scores

We will now introduce you to the data valuation application. Please watch the video below.

If you have problems watching the video, you can also watch it via https://www.youtube.com/watch?v=XptdOe_HcVs

Post-test: Perceptions of application
In this part, we want to understand your perceptions of the data valuation application you just observed.