

## User Perceptions of Multiparty Data Collaboration

You are invited to participate in a research study conducted by Bailey Kacsmar, Kyle Tilbury, and Miti Mazmudar under the supervision of Florian Kerschbaum in the School of Computer Science of the University of Waterloo, Canada. The objectives of the research study are to gather insights into perceptions on multi-party data collaboration.

If you decide to continue, you will be asked to complete an 8-minute online survey comprised of 12 questions that is completed anonymously. It is anonymous in that you are not asked to provide name or other identifying information. Participation in this study is voluntary. You can withdraw your participation at any time by not submitting your responses or exiting the survey. You can choose to not answer any questions that you would feel uncomfortable answering. There are no known or anticipated risks from participating in this study. Upon completion of the study, you will receive remuneration in the amount that you have agreed to with the platform through which you entered this survey. The amount received is taxable. It is your responsibility to report this amount for income tax purposes.

It is important for you to know that any personal information that you provide will be confidential. You may choose to fill in demographic data, such as age-group or gender. Once data is submitted, it cannot be withdrawn as the researchers have no way of knowing which data belongs to which participant. All the data will be summarized, and no individual could be identified from these summarized results. Additionally, anonymous quotations may also be used in any publications that come from this study.

When information is transmitted over the internet confidentiality cannot be guaranteed. University of Waterloo practices are to turn off functions that collect machine identifiers such as IP addresses. SurveyMonkey may collect this information without our knowledge and make this accessible to us. We will not use or save this information at all. If you prefer not to submit your survey responses through this host, please do not participate in this study.

The data, with no personal identifiers, collected from this study will be maintained on a password-protected computer in a restricted access area of the university. As well, the data will be electronically archived after completion of the study and maintained for seven years. The data without identifiers may be shared publicly. Your identity will be confidential.

This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Committee (ORE#41762). If you have questions for the Committee contact the Office of Research Ethics, at 1-519-888-4567 ext. 36005 or ore-ceo@uwaterloo.ca. By agreeing to participate, you are not waiving your legal rights or releasing the investigator(s) or involved institution(s) from their legal and professional responsibilities.

For all other questions about the study, please contact the student researchers: Bailey Kacsmar at bkacsmar@uwaterloo.ca, Kyle Tilbury at ktilbury@uwaterloo.ca, Miti Mazmudar at miti.mazmudar@uwaterloo.ca, or Florian Kerschbaum at florian.kerschbaum@uwaterloo.ca. Further, if you would like to receive a copy of the results of this study, please contact any investigator.

Thank you for considering participation in this study.

### **Consent to Participant**

With full knowledge of all foregoing, I agree, of my own free will, to participate in this study.

[By continuing to the next page] "I agree to participate."

## User Perceptions of Multiparty Data Collaboration

\* 1.

A 8.37% **Scenario:**

TechForYou is a large internet company that offers a search engine, email accounts and smartphone platforms to users. GoodHealth runs a chain of hospitals across the country and stores health data for millions of patients during its day-to-day operations. TechForYou and GoodHealth will share the customer data they hold with one another. You are a customer of TechForYou and GoodHealth.

How acceptable is this scenario?

B 8.33% **Scenario:**

GoodHealth runs a chain of hospitals across the country and stores health data for millions of patients during its day-to-day operations. TechForYou is a large internet company that offers a search engine, email accounts and smartphone platforms to users. GoodHealth and TechForYou will share the customer data they hold with one another. You are a customer of GoodHealth and TechForYou.

How acceptable is this scenario?

C 8.33% **Scenario:**

TechForYou is a large internet company that offers a search engine, email accounts and smartphone platforms to users. GoodHealth runs a chain of hospitals across the country and stores health data for millions of patients during its day-to-day operations. TechForYou and GoodHealth will share the customer data they hold with one another. You are a customer of TechForYou.

How acceptable is this scenario?

D 8.33% **Scenario:**

GoodHealth runs a chain of hospitals across the country and stores health data for millions of patients during its day-to-day operations. TechForYou is a large internet company that offers a search engine, email accounts and smartphone platforms to users. GoodHealth and TechForYou will share the customer data they hold with one another. You are a customer of GoodHealth.

How acceptable is this scenario?

**E 8.33% Scenario:**

TechForYou is a large internet company that offers a search engine, email accounts and smartphone platforms to users. GoodHealth runs a chain of hospitals across the country and stores health data for millions of patients during its day-to-day operations. TechForYou will share its customer data with GoodHealth. You are a customer of TechForYou.

How acceptable is this scenario?

**F 8.33% Scenario:**

GoodHealth runs a chain of hospitals across the country and stores health data for millions of patients during its day-to-day operations. TechForYou is a large internet company that offers a search engine, email accounts and smartphone platforms to users. GoodHealth will share its patients' health records with TechForYou. You are a customer of GoodHealth.

How acceptable is this scenario?

**G 8.33% Scenario:**

TechForYou is a large technology enterprise that wants to acquire additional data about its users. TechForYou wants to buy data about its users from other companies: SmartAd (an online advertising company), MyHomeRetail (a large retail company selling housewares), and MyPurchaseCard (a credit card company). From these other companies, TechForYou will get information on the ads that TechForYou users see, their shopping habits, and their purchasing habits. You are a customer of all these companies.

How acceptable is this scenario?

**H 8.33% Scenario:**

GoodHealth runs a chain of hospitals across the country and stores health data for millions of patients during its day-to-day operations. GoodHealth wants to buy data about its users from other companies: SmartAd (an online advertising company), MyHomeRetail (a large retail company selling housewares), and MyPurchaseCard (a credit card company). From these other companies, GoodHealth will get

information on the ads that GoodHealth users see, their shopping habits, and purchasing habits. You are a customer of all these companies.

How acceptable is this scenario?

I 8.33%

**Scenario:**

NiceStartup is a tech startup that that helps its users track their diet, fitness, and social habits. TechForYou is a large internet company that offers a search engine, email accounts and smartphone platforms to users. TechForYou has bought out NiceStartup and as a result TechForYou now has access to all the data NiceStartup had collected from its users. You are a customer of NiceStartup.

How acceptable is this scenario?

J 8.33%

**Scenario:**

NiceStartup is a tech startup that that helps its users track their diet, fitness, and social habits. GoodHealth runs a chain of hospitals across the country and stores health data for millions of patients during its day-to-day operations. GoodHealth bought out NiceStartup and as a result GoodHealth now has access to all the data NiceStartup had collected from its users. You are a customer of NiceStartup.

How acceptable is this scenario?

K 8.33%

**Scenario:**

AlmostStartup is a tech startup that developed a fitness tracking device. NiceStartup is another tech startup that helps its users track their diet, fitness, and social habits using an app. NiceStartup merged with AlmostStartup. NiceStartup now has access to all the data AlmostStartup had collected from its users.

TechForYou is a large internet company that offers a search engine, email accounts and smartphone platforms to users. Now, TechForYou has bought out NiceStartup and as a result TechForYou now has access to all the data both startups, NiceStartup and AlmostStartup, had collected from their users. You were a customer of AlmostStartup.

How acceptable is this scenario?

L 8.33%

**Scenario:**

AlmostStartup is a tech startup that developed a fitness tracking device. NiceStartup is another tech startup that helps its users track their diet, fitness, and social habits using an app. NiceStartup merged with AlmostStartup. NiceStartup now has access to all the data AlmostStartup had collected from its users.

GoodHealth runs a chain of hospitals across the country and stores health data for millions of patients during its day-to-day operations. Now, GoodHealth has bought out NiceStartup and as a result GoodHealth now has access to all the data both startups, NiceStartup and AlmostStartup, had collected from their users. You were a customer of AlmostStartup.

How acceptable is this scenario?

- ☐ Completely unacceptable    ☐ Somewhat unacceptable    ☐ Neutral    ☐ Somewhat acceptable
- ☐ Completely acceptable

\* 2. How acceptable is the above scenario if the company (or companies) **inform** you by...

	Completely unacceptable	Somewhat unacceptable	Neutral	Somewhat acceptable	Completely acceptable
no formal notification? You learn of the scenario from an independent news report.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
an email or app notification which indicates that by continuing to use their service you agree to your data being shared?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
providing an option to stop your data from being shared? Your data will be shared by default unless you select this option.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
providing an option to allow your data to be shared? Your data will not be shared by default unless you select this option.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

\* 3. How acceptable is the above scenario if the company (or companies) **deletes** the acquired data...

	Completely unacceptable	Somewhat unacceptable	Neutral	Somewhat acceptable	Completely acceptable
never?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
after they have finished using it?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
after a specified length of time (e.g. three months, one year, etc.)?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

\* 4. How acceptable is the above scenario if the company (or companies) **use** the acquired data...

	Completely unacceptable	Somewhat unacceptable	Neutral	Somewhat acceptable	Completely acceptable
to improve the company's (or companies') services?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
for generating advertising revenue?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
in exchange for providing you with a monetary return (e.g. free service, reduced rate, or gift-card)?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

\* 5. What is the shape of a red ball?

☐ Red ☐ Blue ☐ Square ☐ Round

\* 6.

**A 20.0%** The original company will only share approximations of all customers' true data, using what's known as Local Differential Privacy (LDP). LDP guarantees that the odds of correctly guessing your true data from the shared approximation, are low.

How acceptable is the above scenario if the company (or companies) used local differential privacy (LDP)?

**B 20.0%** The new company will create approximations of all customers' true data, using what's known as Central Differential Privacy (CDP). The new company can compute statistics over these approximations, without using your true data anymore. CDP guarantees that the odds of correctly guessing your true data from an approximation, are low.

How acceptable is the above scenario if the company (or companies) used centralized differential privacy (CDP)?

**C 20.0%** Anonymizing the data means that personal identifiers such as your name, age, gender, email and contact address will be removed. Only the anonymized data for all users will be stored together (in a database).

How acceptable is the above scenario if the company (or companies) only shared anonymized data?

**D 20.0%** The company will combine your data with other customers' data and calculate aggregate statistics, such as averages, over all customers' data. The original data will not be shared.

How acceptable is the above scenario if the company (or companies) only shared aggregate statistics?

**E 20.0%** Encryption means means that your data cannot be read by a person or computer who does not have a secret key to remove the encryption. Encryption allows the company to use your data without learning what it actually is.

How acceptable is the above scenario if company (or companies) only shared encrypted data?

- ☐ Completely unacceptable    ☐ Somewhat unacceptable    ☐ Neutral    ☐ Somewhat acceptable
- ☐ Completely acceptable



\* 7. Which of the following best describes the technique in Question 6?

- ☐ Individual data is changed by the original company so that it is hard to later identify the original individuals data contributions.
- ☐ Collected data is received by the new company and they modify the data so that it is hard to identify individual contributions.
- ☐ Parts of the data have been removed so that it is hard to connect to individuals.
- ☐ All users' data is combined and only the results of the computations over the collective data are shared.
- ☐ Individual data is changed such that the original data is not recoverable without additional information.

8. In general, what are your thoughts on companies sharing data with other companies?

## User Perceptions of Multiparty Data Collaboration

\* 9. Please specify your gender:

- ☐ Man
- ☐ Woman
- ☐ Non-binary
- ☐ Prefer not to say
- ☐ I prefer to self describe (please specify)

\* 10. Please select your age group:

- ☐ 18-24
- ☐ 25-34
- ☐ 35-44
- ☐ 45-54
- ☐ 55-64
- ☐ Prefer not to say

\* 11. Which of the following best describes the principal industry of your employment?

\* 12. What is the highest level of school you have completed or the highest degree you have received?

- ☐ Less than high school degree
- ☐ High school degree or equivalent (e.g., GED)
- ☐ Some college but no degree
- ☐ Associate degree
- ☐ Bachelor degree
- ☐ Graduate degree
- ☐ Completed college (other degree or certificate)
- ☐ Prefer not to say

## User Perceptions of Multiparty Data Collaboration

Dear Participant,

We would like to thank you for your participation in this study on user perceptions of multiparty data collaboration. As a reminder, the purpose of this study is to understand people's intentions to consent to multiparty data collaboration. We are investigating people's perceptions of separate companies pooling their collective data directly as well as their perceptions of such companies using some privacy preserving techniques to protect individual's data. We are also investigating whether the application that the data is being used for, has an impact on perceptions. This study will advance researchers' and policy makers' understanding of people's preferences on multiple companies or parties sharing their data among themselves. In turn, this will help us and other researchers design privacy-enhancing technologies that protect people's privacy, while sharing their data. We did not disclose that this study was also about privacy at the start of the survey. We withheld this information to avoid biasing your responses. We apologize for not providing full information at the start of the survey.

All participants will be able to access a summary of the findings of the study in January 2022 by contacting the student researchers (contact info below). The results will be shared through publication at an academic venue such as a conference or journal.

Data without identifiers may be shared publicly. Your identity will be confidential. Your data will be securely stored on a password-protected computer in a restricted access area of the university. Your response is anonymous in that no identifying information was collected or will be used in the presentation of results. The data will be retained for a minimum of 7 years after the conclusion of the study.

This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Committee (ORE#41762). If you have questions for the Committee contact the Office of Research Ethics, at 1-519-888-4567 ext. 36005 or ore-ceo@uwaterloo.ca. For all other questions contact the student researchers: Bailey Kacsma at bkacsma@uwaterloo.ca, Kyle Tilbury at ktilbury@uwaterloo.ca, or Miti Mazmudar at miti.mazmudar@uwaterloo.ca.

Thank you again,  
Bailey Kacsma, Kyle Tilbury, and Miti Mazmudar