D3.3 User Studies Analysis

Mads Schaarup Andersen(ALX)

The project SODA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731583.

Project Information

Scalable Oblivious Data Analytics

Project number:	731583
Strategic objective:	H2020-ICT-2016-1
Starting date:	2017-01-01
Ending date:	2019-12-31
Website:	https://www.soda-project.eu/

Document Information

Title: User Studies Analysis

ID:	D3.3	Type:	R	Dissemination level:	PU
		Month:	M36	Release date:	31.December.2019

Contributors, Editor & Reviewer Information

Contributors (person(partner): sections)	Mads Schaarup Andersen(ALX): All Sections
Editor (person/partner)	Mads Schaarup Andersen(ALX)
Reviewer (person/partner)	Peter Scholl(AU)

Release History

Release number	Date issued	SVN version	Release description / changes made
1.0	December 18, 2019	NA	Initial version
1.1	December 23, 2019	NA	Changes made based on internal review
1.2	December 24, 2019	NA	Editor added and minor formatting changes

SODA Consortium

Full Name	Abbreviated Name	Country
Philips Electronics Nederland B.V.	PHI	Netherlands
Alexandra Institute	ALX	Denmark
Aarhus University	AU	Denmark
Göttingen University	GU	Germany
Eindhoven University of Technology	TUE	Netherlands

Table 1: Consortium Members

1 Executive summary

This deliverable documents the user studies and the analysis of these. This is done by going through each of the activities carried out as part of the user studies in the SODA project. The user studies are divided into the two user groups: end users, or data analysts, and data subjects, or patients. The latter group is further divided into people with dementia and people with diabetes. Finally, the main results are described. The user studies were carried out with a goal to understand why MPC is not wider applied, and what can be done to help the further adaptation of MPC technology. We have chosen to explore MPC mainly from a big data perspective in the health sector, i.e., how can MPC be used in big data analysis in the health sector and how do we make patients aware of how their data is being used.

We find that MPC has potential uses in the health sector, but that there are also a number of obstacles. These obstacles are because of the properties of MPC such as data visibility and data leakage when querying the same data several times. However, we also found that there are a number of non-MPC related issues in data sharing in the health system today which need to be understood in order to make decisions about whether or not to setup an MPC system. Identifying and understanding these issues can, furthermore, help make better use of data in the Danish and UK health sectors with but also without MPC. The non-MPC related issues include organizational and legal issues as well as a lack of knowledge. We find that instead of a technology problem, often what is needed in setting up MPC collaborations is deep knowledge of data sharing and a process around setting up MPC collaborations. We propose a process for carrying out MPC collaboration workshops and test it out.

Next, we find that data subjects, or patients, have different data sharing preferences depending on their condition, so they cannot necessarily be treated as a homogeneous group. This is based on our user studies in which we find that the needs of people with dementia in some ways are different from people with diabetes. However, common for the groups studied which might not be similar to other types of patients is that both groups do not consider themselves patients, but rather people who have some condition which is part of their lives. This has implications in how consent interfaces should be designed differently for different groups and we propose suggestions on how to design them.

Finally, we find that not having a common language to talk about trust might prevent MPC from being adopted more widely. This is based on how the MPC community typically talks about trust as something to be solved technically and mathematically, whereas data subjects and end users in our studies think more of trust in terms of organizations and institutions. This difference in communication might make it difficult to foster a conversation about MPC in which it is clear exactly what the technology offers. We suggest that a common language to talk about trust and MPC is created.

2 About this Document

2.1 Role of the deliverable

This deliverable presents the analysis of the user studies conducted as part of the SODA project.

2.2 Relationship to other SODA deliverables

This deliverable is related to the work on the demonstrators in WP4. In particular the results presented in this deliverable have been used to inform some of the work presented in deliverable D4.5.

2.3 Relationship to other versions of this deliverable

N/A

2.4 Structure of this document

The document is structured in the following way. Section 4 is the introduction, Section 5 contains the description of activities involved in the study of the end users, Section 6 contains the description of activities involved in the study of the data subjects, and finally, Section 7 contains a description of the main contributions from the user studies.

3 Table of Contents

1 Executive summary **7**

2 About this Document **8**
2.1 Role of the deliverable . . . 8
2.2 Relationship to other SODA deliverables . . . 8
2.3 Relationship to other versions of this deliverable . . . 8
2.4 Structure of this document . . . 8

3 Table of Contents **9**

4 Introduction **10**
4.1 User Studies in SODA . . . 10
4.2 Different Settings and Subcontractors . . . 10
4.3 Overall Approach . . . 10

5 End Users **11**
5.1 Interviews . . . 11
5.2 Workshops . . . 13
5.2.1 Workshop 1: Explorative Workshop on Data Sharing . . . 13
5.2.2 Workshop 2: Introducing MPC Workshop . . . 14
5.2.3 Workshop 3: Setting up an MPC collaboration . . . 18

6 Data Subjects **22**
6.1 Diabetes . . . 22
6.1.1 Workshop 1: Explorative Workshop on Data Sharing . . . 22
6.1.2 Workshop 2: MPC and Data Sharing . . . 23
6.1.3 Video and Interviews . . . 24
6.1.4 Phone Interviews . . . 25
6.2 Dementia . . . 25
6.2.1 Explorative Workshop . . . 25
6.2.2 Three Workshop Sessions . . . 25

7 Main Results **27**
7.1 Applying MPC to Real World Problems . . . 27
7.2 MPC and Data Subjects . . . 28
7.3 Trust is not equal to Trust? . . . 30

References **31**

4 Introduction

Secure Multi-Party Computation (MPC) as a technology, has existed as an idea for more than 30 years [11], however it was not until 2009 [9] that we saw the first real world implementation of an MPC system, which was the Danish sugar beet secret auctioning system. Since then, we have only seen a few real-world implementations such as the Boston pay gap study [12] and there are still only a few companies out there that provide MPC solutions such as: Unbound Tech [8], Partisia [5], Sepior [7], Inpher [4], AKEYLESS [1], and Cybernetica [2] - most of which apply MPC to very well-defined problems such as key-management and secret auctions. At the same time MPC is mostly presented as a general technology that can solve a lot of existing problems in a secure and private way. This begs to ask to the question as to why more companies are not using MPC and why the technology is not more widely applied. To answer that question, we need to approach the subject from a user centric point of view. To explore this, part of purpose of the SODA project is to explore this question from a patient, or data subject, point of view as well as from the data analyst, or end user, point of view. In this deliverable, we will describe each activity of the user studies of SODA along with the main results from the activity. Activities are divided into a section on end users and one on data subjects. The data subject section is subdivided into people with dementia and people with diabetes. Finally, we report the main findings of the user studies.

4.1 User Studies in SODA

In this project we have conducted user studies on MPC within the medical domain. We have worked with two types of users within the domain: 1) data subjects, which are users that provide their data to some sort of data analysis and 2) end users, which are users that either conduct data analysis or rely on such an analysis.

4.2 Different Settings and Subcontractors

One of the overall goals of the project is to improve the privacy of data subjects which in the case of SODA are patients. Since privacy is a subjective thing which is influenced by culture, it is interesting to look at different types of medical patients and different cultures. Therefore, it was chosen to study people with dementia and diabetics (pregnant and non-pregnant) and to study two different cultures. To do this, two subcontractors were found to help facilitate the contact to patients. The choice ended being Lancaster University to help with dementia patients in a UK context and Aarhus University Hospital (AUH) to help with diabetes patients in a Danish context.

4.3 Overall Approach

We have taken a qualitative approach in which we have used semi-structured interviews and different types of workshops. We have chosen this approach as these methods allow us to be explorative - something which is important in this unexplored area of MPC research.

5 End Users

The studies with the end users is motivated by a desire to get a better understanding of, from a high level perspective, why MPC is not more widely used. The idea from the beginning of the project was to use this understanding to design a tool or a plugin for an existing data analysis tool such as SPSS [3] or SAS [6]. This was motivated by a desire to create a user experience close to the existing one with minor tweaks due to the properties of MPC data sharing. However, during our interviews, it quickly became clear that there are a lot of underlying problems in data sharing in the Danish and UK health sector that need to be understood.

5.1 Interviews

Purpose: The purpose of the interviews was to get deep knowledge of how data is used and shared within the Danish health sector with a focus on barriers and opportunities. Furthermore, the purpose was to get the interviewees' perspective on how MPC technology could potentially benefit them and what barriers MPC technology would present.

Participants: 13 interviews were conducted with the following stakeholders from the Danish health sector.

- Steno Diabetes Center, Aarhus University Hospital
- Center for Cancer Research, Danish Cancer Society
- Business Intelligence Unit, Region Midt
- Clinical Epidemiological Unit, Aarhus University Hospital
- Danish Health Data Authority (Software Architect)
- Danish Health Data Authority (Project Manager)
- Data Unit, Aarhus University Hospital
- Pre-hospital Unit, Aarhus Municipality
- RKKP - Quality Control Program of The Danish Regions
- KiAP - Quality in General Practice
- Department of Public Health, Aarhus University
- Novax, a private company
- Enversion, a privacy company

Method: The interviews were carried out as qualitative, semi-structured interviews lasting around one hour each. The interviews were conducted using an interview guide adapted to specific stakeholders, but all the interviews were structured using the following format:

1. **Organization** Questions about the organization and the job of the interviewee
2. **Data** Questions about what types of data the organization collects, uses, and holds
3. **Data Use and Sharing** Questions about how data is used, shared, and how/whether data from other sources are used
4. **Data Security and Privacy** Questions about the sensitivity of the data, how this is treated, and what kind of security and privacy measures are being applied
5. **Secure Multi Party Computation** After getting a very brief explanation, the interviewee was asked about potentials and possible barriers of using MPC in their setup

Two researchers, a computer scientist and an anthropologist, participated in all interviews and both wrote notes. These notes were analyzed using qualitative coding.

The participants were selected in collaboration with AUH to represent the Danish health sector broadly in areas where data is a large part of the work.

Analysis:

There is a lot of focus on data There is currently a lot of focus on data and using data in the Danish health sector. Participants mentioned the following areas as potentials for more data use: Quality development, medical treatment, efficiency improvement, benchmarking, reporting, invoicing, learning, creating dialogue, operation, measurement, development, feedback, screening, uniformity. They also mentioned the following challenges: Law, GDPR, surveillance, leakage, legal basis of processing, access, ownership, increasing amounts of data, standardization, resource consumption.

Data is not just data Data is not objective and neutral. Data is collected in a specific context with a purpose. Reusing data requires knowledge of how and for what purpose data was originally collected. For example, an interviewee mentioned that self reported activity data from smart phones has the problem that men typically carry their phone in the pocket, whereas women typically carry it in a purse leading to different measurements for the same activity. Another participant mentioned that in one of their systems the unit for measuring weight changed from grams to kilograms when a child turned two years old. This lead to some interesting problems when a child suddenly appeared to weight several tons. Currently, error handling and correcting for missing data is part of the data management process.

Homogeneous data requires a lot of work The process of combining data today, usually involves a lot of data management and cleaning. This is usually done by collecting all the data in one place. The process involves, e.g., making sure that data fields are called the same for the same data and making sure that the unit is the same (e.g., gram vs. kilograms). When only partial data is present, there also needs to be made a decision of whether or not to include it in the analysis and this also requires data in one place. Furthermore, there is sometimes a need to know something about the data in the process of deciding whether to include it or not, i.e., some meta-data might be needed to make

this decision. Finally, a lot of energy and time is also spent on physically moving large amounts of data around.

Barriers for using MPC Some participants mentioned that their work involved an explorative process where an analysis was run on the data, then reevaluated and run again, for a, from the start, unknown amount of iterations. Hence, in this type of problems, it is not known up front exactly which analysis will be run on the data when collecting from different sources. Hence, this explorative way is not necessarily supported by an MPC system. This is a problem in relation to MPC as running multiple custom queries can leak a lot of the data if done in a clever way. Some interviewees also expressed the need to have access to all the data, since visual inspection was needed to either argue for a medical treatment or to publish research based on the data. Finally, one interviewee also mentioned that the GDPR would sometimes stand in the way, as data controllers were afraid to do something wrong, and would say no to allow for data access just to be on the sure side.

Potentials for MPC There was a few participants that mentioned that the things they wanted to calculate was known up front such as benchmarking clusters of general practitioners against others or getting a specific number calculated at a hospital. This makes a good case for MPC. Others suggested that obtaining data from multiple sources was a very tedious process and sometimes they would end up with data they could not use anyway. They suggested that MPC could be used for an initial hypothesis testing which could then inform whether or not setting the data gathering in motion was worth the effort. One interviewee also mentioned that if MPC could help eliminate the need to move data around, it would be very useful.

5.2 Workshops

A total of three workshops were conducted with the end users. Two in the UK and one in Denmark. The workshops were structured so that they build on results from the previous workshops. The UK workshops were conducted prior to the one in Denmark. Furthermore, the workshops in the UK were structured around people suffering from dementia, whereas the workshop in Denmark was conducted with a broader topic as

5.2.1 Workshop 1: Explorative Workshop on Data Sharing

Purpose: Explore the main issues, concerns, questions, feelings, and perceptions of using technologies for sharing data when living with dementia. This workshop was also used as a kick-off for all the UK and dementia related activities. Therefore, both end users and data subjects were invited. Hence, it serves as input to insights about both the data subjects and the end users. The workshop was conducted in Lancaster.

Participants: The participants were recruited using contacts from within Lancaster University and organizations in the UK working with Dementia patients. Participants were recruited using email. There were 13 participants at the workshop. Participation was anonymous, but the following groups were represented:

- Lancaster University Experts
- Professional Caregivers

- People with early stage dementia
- Family or Individual Caregivers
- Data scientists and managers

Method: The workshop was organized in three phases: First, participants had to brainstorm the following questions, writing each question down: What is your data? What is the data you need? How do you feel about sharing data? How do you imagine data? Who owns your data? What is personal and what is medical data? What are the benefits of data sharing? What are your doubts about data sharing? Second, participants were divided into smaller groups where the answers were classified and organized. This was done using post-its and the results from one groups can be seen in Figure 1. Third, based on the categorization, the categories were put on boxes - the idea being that by placing a box on the floor, the had to eliminate the category they found the least important. An example of this can be seen in Figure 2.

Analysis:

I am not a patient, I am a person Participants expressed the need to be viewed as persons rather than patients. One participant, a person with early stage dementia, expressed a feel of being left out of conversations after she was diagnosed, exemplified with a situation where someone had asked her carer: "Does he take sugar?" talking directly to the carer, disregarding that the participant was even in the room. This left her with a feeling of being her condition rather than a person.

Addressing vulnerable groups Addressing vulnerable groups can be challenging. Co-creation and activities must be run in a safe and comfortable environment, many times stories emerge that are of sensible nature and conversations are difficult to hold, because it is emotionally charged. This is important for the following activities with people with dementia.

Data sharing Data sharing was discussed among the participants and one of the main things that the participants described was a desire for the data to be used for good or for improving treatment.

5.2.2 Workshop 2: Introducing MPC Workshop

Purpose: Map out real data flows in the UK health system and using these data flows to facilitate a session to explore how MPC might be utilized. The workshop took place in Lancaster.

Participants: Participants were recruited through Lancaster University contacts and a recruitment email was sent out. The workshop had a total of 12 participants + the researchers. The specific participants have been anonymized and the following groups were represented:

- Lancaster University experts
- NHS clinicians
- Academic and non-academic medical data scientists
- NHS business intelligence

Figure 1: Results of the discussions among one of the groups.

- Data scientists and managers
- Data regulatory body officers
- Lancaster city council information office officers

Method: The workshop was a full day in which it started out by mapping out data flows in the UK health sector. During the session five researchers participated. three university researchers from Lancaster University and two from ALX. The workshop was conducted at Lancaster University, and

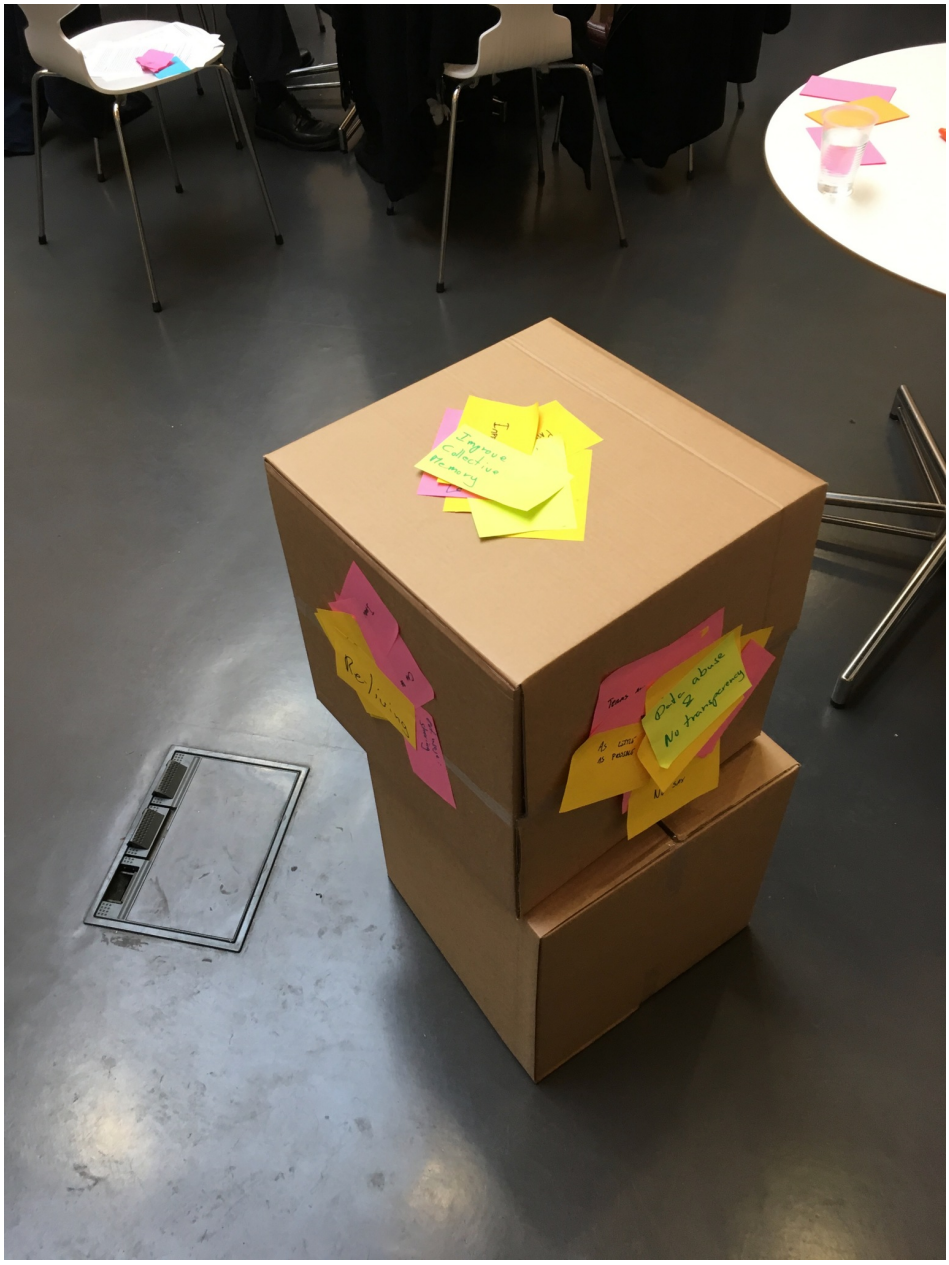

Figure 2: Results of one of the groups work with boxes.

received approval from the local ethics board.

The workshop was divided in three parts. First, there was a presentation of the purpose of the project along with a short presentation on what MPC technology is and how it is used. Second, there was an interactive session in which all participants placed figures and drew on a table while discussing which data goes where. The result can be seen in Figure 3. Third, participants were divided into smaller groups where they discussed how MPC might be applied. During the second phase, one researcher facilitated while the other four listened to the conversation and took notes. In addition, audio recorders where place in the room to record the conversation.

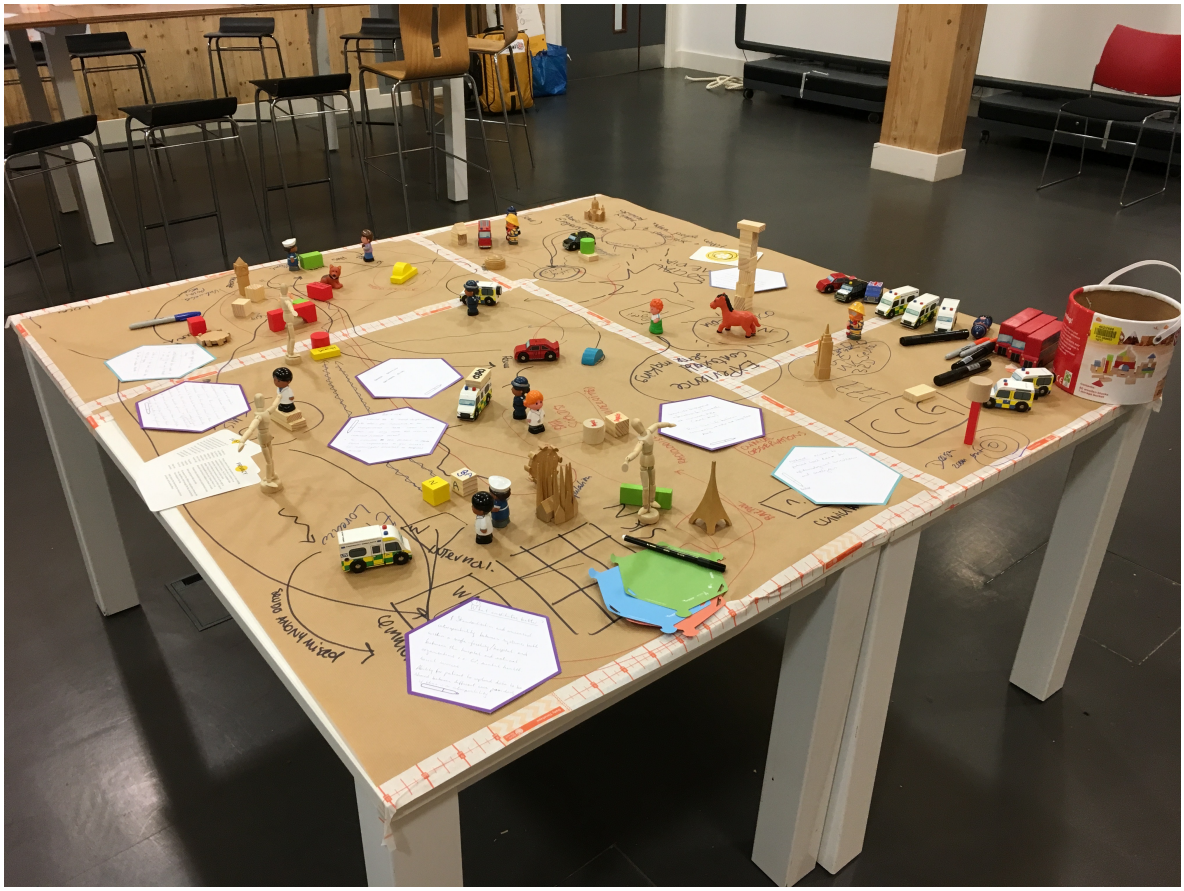

Figure 3: Results of the mapping of data flows.

Analysis:

MPC can be useful During the workshop it became clear that MPC technology could be a useful way to combine data from different sources in situations where it might be impossible of difficult to do this without MPC technology due to legal or practical issues.

Understanding MPC is difficult MPC can be difficult to explain at the level of detail that is needed to fully grasp what is going on. This became clear during the workshop, both during phase 1 where the MPC expert had to explain MPC and a couple of times correct participants when they talked about what they wanted to use MPC for. This also became clear in the third phase where participants were discussing how to apply MPC technology to the UK health sector. During these sessions, the researchers had to pitch in to explain exactly what MPC could and could not do. One group even came up with the concept of introducing a trusted third party to set up MPC system – even though eliminating the trusted third party was party of the concept they had to discuss.

You cannot eliminate trust MPC is often said to be a technology that eliminates the need for a trusted third party. This was also the way it was presented to the participants during the workshop. However, discussions with end-users speculated on the introduction of a certified trusted party such as

NHS or Public Health England to act as arbiter and translator, policing and administrating the MPC.

However, several participants questioned why it would be necessary to use such a technology within the NHS. This was based on that the NHS is normally seen as a trusted third party, leaving no need to be eliminated from data analysis. These participants also indicated that it might have unwanted consequences not foster a distrust in the NHS. Hence, it is not necessarily always a good thing to eliminate trust as trust can be an enabler for making things happen.

Process around MPC collaborations is needed A premise of MPC is that all parties involved agreed on what the system can compute – which questions can be "asked". However, the process of deciding questions, how they should be asked, how data is collected and managed to support beneficial use of MPC techniques, present issues of governability and system stewardship. It was clear from the workshop that it is very difficult to understand exactly what is needed to set up an MPC collaboration. Having the participants discuss in smaller groups made it clear that rather than an explanation of what MPC can do and what it is, there needs to be some process around setting up an MPC collaboration. The discussions raised all sorts of different questions: How often could questions change? What happens to old questions and answers? What if someone wants to leave or join the system? Will that affect guidelines for questions and data input? How trusted arbiter's transparency and accountability are guarantee? Who is allowed to ask what questions? What is allowed to do what with results?

Towards an MPC system Furthermore, end-user discussions evidenced a need for structuring a system that afford a widest range of questions, the system cannot be an "oracle", as every question asked will leak information about the combined data sets, introducing then issues of privacy, purpose binding, and data minimization. Questions above show structuring an MPC system in the health domain presents interoperability challenges but also design opportunities. They also show a need for sufficient information about MPC and its limitations. The conversation about the introduction of trusted third party, partly would defeat one of the main purposes of MPC: to avoid trusting a third party. The workshops show that participants may not trust non-human governability of sensible data.

5.2.3 Workshop 3: Setting up an MPC collaboration

Purpose: The purpose of this workshop was to get different stakeholders from the Danish health sector together and go through the process of setting up an MPC collaboration and, hence, test out a workshop format as a mediator of an MPC discussion. The workshop was based on the output of workshop 2 (Section 5.2.2).

Participants: The participants were recruited among the interviewees and their organizations. 4 participants responded and participated in the workshop. The participants were from the following organizations:

- Novax, a private company
- Department of Public Health, Aarhus University
- Data Unit, Aarhus University Hospital
- Clinical Epidemiological Unit, Aarhus University Hospital

Apart from the participants, 3 researchers participated, two with expert MPC knowledge and one facilitator.

Method: The workshop was a half day workshop divided into three phases: First, the preliminary results from the project were presented followed by a presentation of MPC technology and use. Second, a phase where the participant had to discuss what they wanted/needed the system to calculate and which data would be needed in order to do this, was facilitated as an exercise in which the participants had to write on the table staring with the overall problem to solve, next the specific questions to ask the system, ending with the data sources needed. The table from the workshop can be seen in Figure 4. This was done to explore the discussions around setting up an MPC collaboration. Third, participants had to physically map out data flows in order to think about the specifics of what data would move around and what would be needed to make this possible. The results of this can be seen in Figure 5.

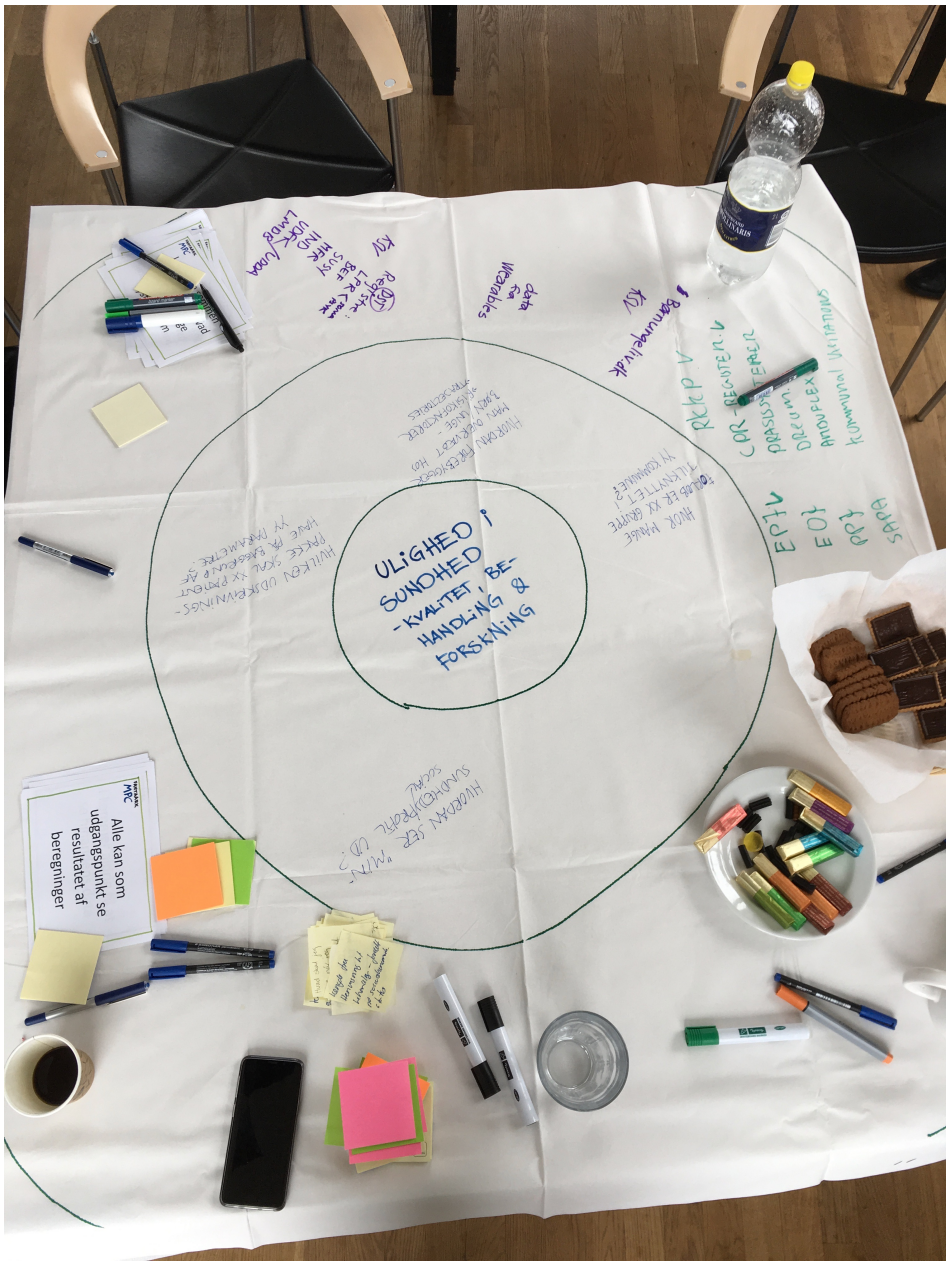

Figure 4: Input for the MPC discussion drawn on a table.

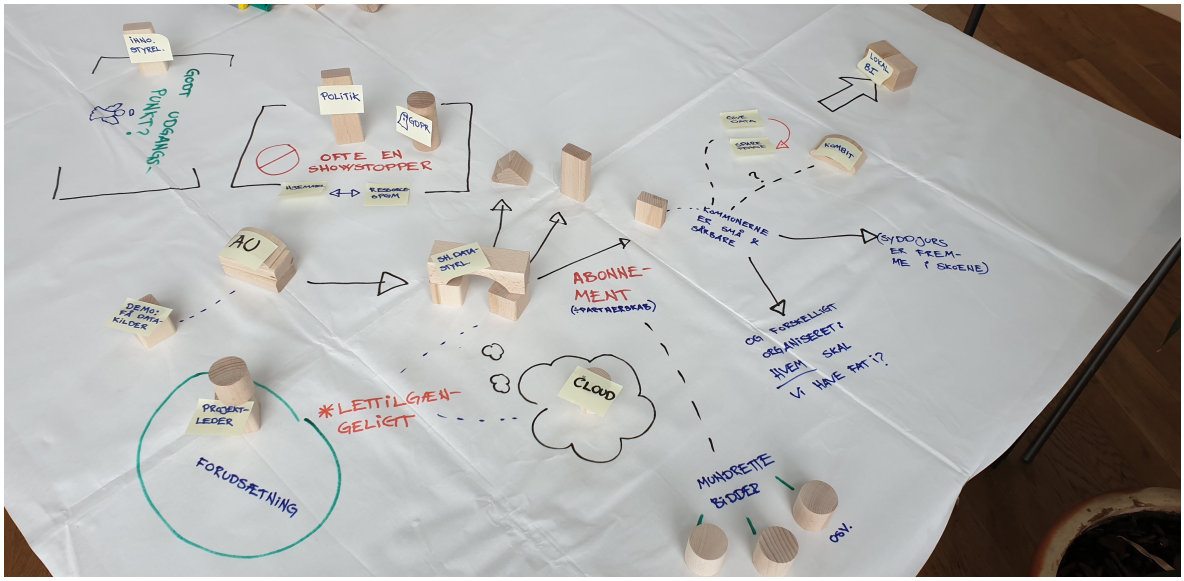

Figure 5: The proposed MPC system with obstacles and barriers mapped out on a table.

As part of the discussion in phase two and three, the participants were provided with an MPC fact sheet with the following statments to help facilitate the discussion (translated from Danish):

- The other participants will never see your data
- All need to agree on what to ask the system
- The system is black box. No-one can visually inspect the data that is not located within your own organization
- Data will only leave your servers in unreadable form and will never exist with other participants in a readable format
- As a rule, all participants can see the result of the calculation
- Data will never be collected in one place

Analysis:

Potentials in MPC During the workshop, it became clear that the participants found that there would be some interesting areas in which they could benefit from combining data which is not possible today. The participants chose to focus on the overall problem of "inequality in health". All suggested uses of MPC were based on getting access to and insights from data sources which can not be combined in the current setup. There was a focus on getting statistical data as well as data on the individual level. Statistical measurements were focused on saying something about how a certain group were doing or which kind of data influenced, e.g., whether a person was re-admitted to the hospital or what the risk factors are in child obesity. The individual measure were focused on being able for a patient to get an individual health profile and what "package" to assign to a patient. Some of these things are possible today, but the participants focused on the fact that getting more data involved in the computations could lead to better decisions. It was suggested that to get the ball rolling, setting

up test systems might be a way to get people on board, but that it is also very important to make it easily understandable.

Challenges in setting up MPC systems During the workshop, several challenges in setting up MPC also emerged. There was a discussion of where a system like this would "live", i.e, where it would physically be installed. It was proposed that it should exist at the universities, however, it was also mentioned that this might be a problem for private companies. It was agreed that it should be some form of public authority when dealing with health data as private companies might not have the legal capacity or be trusted enough, something which is also supported by the results from one of the workshops with diabetics (Section 6.1.2). Incentives for participating were also discussed as a system would be costly to setup and maintain. A subscription service was proposed. Another problem of data sharing mentioned, was that three silos exist in the Danish system today. The first is the electronic patient journal, which contains hospital data, which resides in the regional domain. The second is data from general practitioners. The third is data which resides as the municipality. All these sources of data are relevant in health analysis, but often the three different parties suffer from an unawareness of what data resides in the other silos - that is, if they are even aware that relevant data exists.

Workshops as a Mediator for MPC Discussion Overall, we found that the workshop format worked well in terms of getting a discussion around how MPC could be used. With the help of an initial explanation and the MPC fact sheets during the discussions, the participants were able to discuss how to combine data within the limits imposed by MPC technology. However, during the sessions, participants would also ask the MPC experts, so a recommendation would also to have an expert present when discussing.

6 Data Subjects

6.1 Diabetes

6.1.1 Workshop 1: Explorative Workshop on Data Sharing

Purpose: The purpose of this workshop was to explore what data exists about people with diabetes, who this information is distributed to, and what the sharing preferences of the group are.

Participants: Participants were recruited using a poster in the waiting room at the diabetes unit at AUH and using a members mailing list at the Danish Cancer Society. 8 participants showed up. Three of those were type 1 diabetics and five were type 2.

Method: The workshop was organized in two parts: First, a plenum session in which the project was shortly introduced and an initial brainstorm. Second, the patients were divided into groups by their diabetes types. In these groups they then had to do a more focused brain storm in which they had to identify: 1) the types of data they generate and use in relation to their condition, 2) the ways in which this data is collected, and 3) who will receive this data. This was done on post-its, and the results from one of the two groups can be seen in Figure 6.

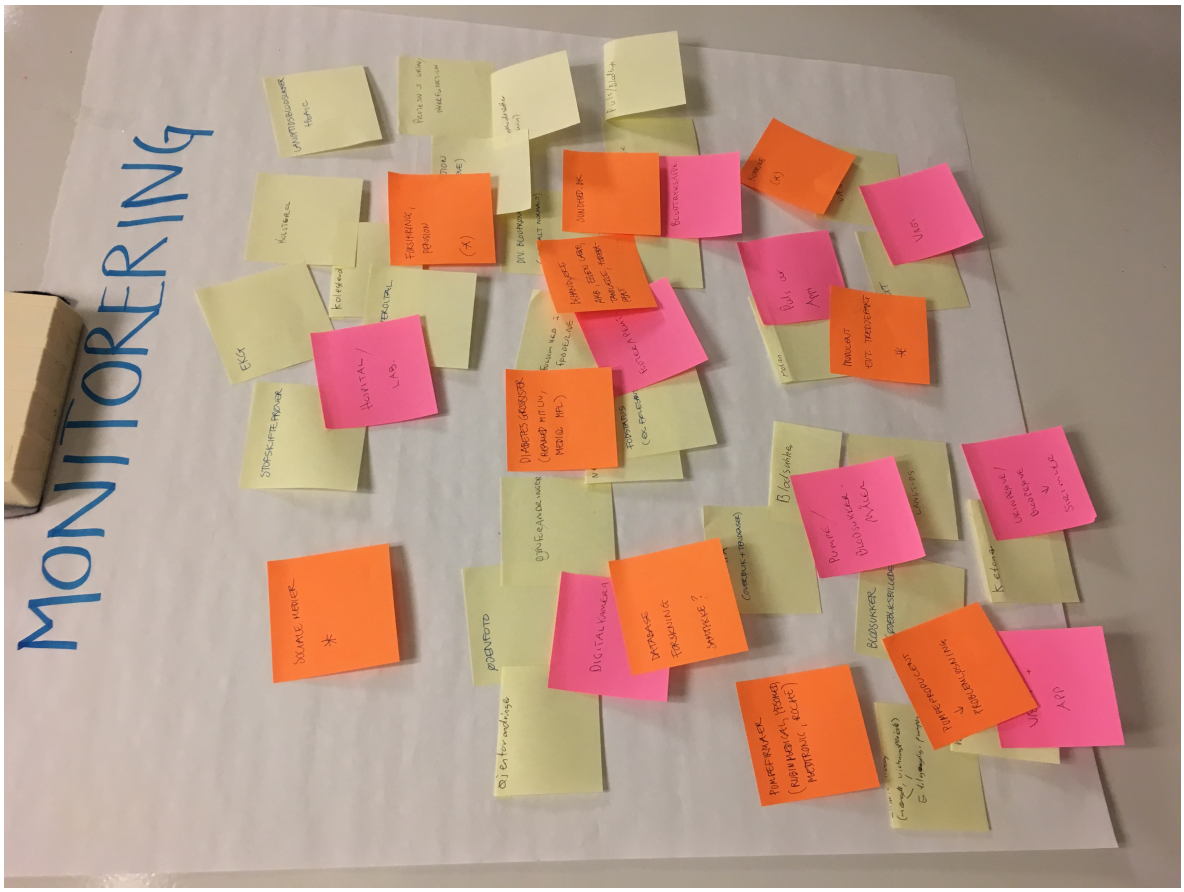

Figure 6: Results of the discussions among the type 1 diabetics.

Analysis:

Focus on helping yourself and others The participants were focused on how they might benefit from sharing more data. Moreover, there was also a willingness to help others who suffered from diabetes. For example, a participant expressed that she would sometimes share pictures of where she would place a device, on Facebook so that others could see what worked for her. Participants expressed using Danish Cancer Society events as well as Facebook groups to share information with other diabetics and get information.

Privacy is important for diabetics During the workshop, the participants expressed different privacy concerns. There were mainly two groups that the participants were very aware not to share data about their condition with: 1) Insurance companies, as it might change the premium on some of their insurances and 2) employers as one participants put it, awareness of diabetes might make an employer select another candidate for a job. However, participants were willing to share data if it might help other diabetics or if they would gain something from sharing data.

Type 1 and type 2 diabetics are different As several participants noted, type 1 and type 2 diabetics suffer from two different conditions. This was very important to many of the participants. Three even stated their type when signing up, even though this was not a requirement. During the plenum session it was also very clear that the groups saw themselves as distinct and they more than once referred to the difference as an "us and them"-thing with statements like: "but of course it if different for *them*".

6.1.2 Workshop 2: MPC and Data Sharing

Purpose: The purpose of this workshop was to explore whether MPC could be understood, scenarios in a diabetes context in which MPC might be applied, and how trust plays a role when sharing data with and without MPC.

Participants: Participants were recruited using the Danish Cancer Society Aarhus mailing list. A total of 9 participants showed up.

Method: The workshop lasted 3 hours and was structured in three phases: 1) A round table discussion about who gets the data, 2) an explanation of MPC technology, and 3) a round table discussion about how what MPC might be used for in a diabetes context and what role trust plays in this scenario.

Analysis:

Trust is complex As part of the workshop, the participants had to discuss which organizations they trusted. While a few participants remained skeptical that anyone could be trusted fully, the overall impression was that the organizations getting data from the diabetics could be divided into the following two categories:

- **High trust** General practitioner, the public health sector, Diasend [1]
- **Low trust** Insurance companies, medical industry

[1] Diasend is a software application used in the public health sector for diabetics to send in measurements.

Trust is placed in organizations rather than technology When asked whether the trusted the technology based on the explanation given, the participants expressed that they needed some entity in which to place their trust. There was a need for some organization to vouch for the MPC system. When asked who that might be, the participants had a long discussion, but found it a bit hard to find a entity that they could agree on. The discussion ended by the participants agreeing that if the EU would vouch for the technology, they would trust it.

Data sharing can create a sense of community A participant expressed that data sharing might be a mediator in creating a sense of community among diabetics. This confirms the finding from the workshop 1, in which participants stated that they used Facebook to gain and share information that might benefit themselves and other diabetics.

There is potential for MPC use The third phase of the workshop was about possible uses for MPC. The participants ended up with two scenarios in which MPC could be used. Th first was a system in which they could benchmark themselves against other diabetics with similar characteristics. This might help in making decisions about when to take insulin, what to eat, and when to exercise. This would be based on all the data available about them at different sources, e.g., general practitioner, eye doctor, foot specialist, medical device companies, their own devices (such as Fitbits), hospital data, etc. The other was a system in which they could get specific advice on whether or not to eat something specific. This would be again be based on all the data that exists about themselves and other at different sources, but specifically it would return a recommendation on whether or not to eat something.

6.1.3 Video and Interviews

Purpose: The purpose of this activity was to develop a video based on the results from the two workshops with diabetics and try to explain MPC technology to them.

Participants: Participants were recruited in the waiting room of the diabetes unit at Aarhus University hospital. Participation was voluntary and anonymous.

Method: A video was created using Videoscribe explaining MPC to data subjects. The video was tested by showing the video to diabetics followed up by asking them a number of questions.

Analysis:

Basic principles of MPC can be explained Based on the video, the participants had to answer high level questions regarding MPC. This showed that the basic principles of MPC can be explained to data subjects.

Understanding MPC is not the main concern While the participants understood the basic principles of MPC, it was not the main concern expressed. The participants were more interested in what they might gain from MPC technology.

6.1.4 Phone Interviews

Purpose: The purpose of the phone interviews was to gather knowledge on pregnant diabetics data sharing practices and preferences.

Participants: Participants were recruited by the diabetics unit at AUH. Participants were given physical recruitment material when they came in for a check up. The recruitment material included a phone number and email address of one of the researchers. A total of 10 participants were recruited and interviewed over the phone.

Method: The interviews were conducted as semi-structured phone interviews.

Analysis:

Many contexts A pregnant diabetic is also a girlfriend, employee, citizen, etc. All the contexts influence how she relates to data sharing. If she, e.g., has come into contact with data security through her work, this influences her data sharing as a pregnant diabetic.

Common experiences The diabetics interviewed had different experiences and views on things, but they all had the following four things in common: 1) being a pregnant diabetic is hard work, 2) blood sugar levels are a constant worry, 3) sharing data is seen as a requirement for receiving qualified guidance, and 4) there is a lot of trust of health personnel that deals with diabetes.

Motivators for sharing data The interviewees expressed the three following reasons for sharing data: 1) to help research, 2) be a role model for other pregnant diabetics, and 3) sharing data is a prerequisite for getting qualified guidance in their condition.

6.2 Dementia

Part of the work with dementia patients was published at IASDR 2019 [10].

6.2.1 Explorative Workshop

As mentioned above, the first workshop held in the UK served to gain insights about both data subjects and end users and, hence, the workshop is described in Section 5.2.1

6.2.2 Three Workshop Sessions

Purpose: The purpose of the sessions was to gather knowledge on people with dementia and how to do research with this vulnable group.

Participants: Participants were recruited through contact with NeuroDropin Center. A total of six participants living with dementia and two caregivers participated in the three workshops.

Method: The first and second sessions were dedicated to the co-creation of Persona and Scenarios respectively. Personas captured conversations about objects, stories, things and numbers that matter for individual data subjects living with dementia. The third session regarded participants as experts rather than patients, as per the results from the explorative workshop. This was essential for the sessions to run effectively. The last tool to facilitate co-design was used in a dynamic that emulated *an expert interview*. Answers were recorded visually on the wall. Participants expressed pride and fulfillment at their voices being taken seriously. Each of the sessions lasted around 2 hours.

Analysis:

Proportionality of data usage Participants expressed concern about how the data usage in an MPC system might benefit them. As one participant expressed it: "If during my everyday activities such as collecting my prescription at the pharmacy, visiting doctors and care center, I am sharing information, such as my name and address with so many agencies and organizations, because my data is may be useful for them, how is it that MPC cannot be designed so my data is also useful to me? I usually forget my own address. How could this MPC work both ways? How could MPC be useful to me?"

Transparency, fairness, and justice The process of retrieving information (asking questions) to an MPC system is very opaque. How can data be designed to be accessible, palpable, and to be just and inclusive. Can data systems help people living with dementia to remember their addresses and personal information? How can data systems be designed to benefit data subjects in fair data transactions? How can the data subject's interests, values, and needs be preserved in the design of operation of data systems?

Co-designing for people with dementia The project designed and re-purposed new and existing co-design tools and activities that facilitated the co-design process informed by the specificities of SODA contexts and participants, and allowed us to develop a deeper understanding of data subjects' perceptions of their own data sharing and how it influences their wellbeing.

Research with vulnerable people Looking for the lost vulnerable data subject in the complex obscure big data system, was a way to respond to the dignity call. The experience of co-designing ethics with people that are feeling unwell and whose health is deteriorating quickly, make of the space of affections that co-design opens, a very fragile one, that call for care, for care informing research practices and care informing relationships. We propose a process of technology innovation fueled by ethics, dependent upon trust that emerges and exercises within the fragility of creative encounters. It is simultaneously about staying responsible with the co-design process, loyal to its subjects and matters, and about taking care of one's abilities to respond to it.

7 Main Results

As mentioned in the introduction, one of the main goals of the user studies was to try to answer the question of why MPC is not more widely adopted. During the user studies, we have explored this from both a data subject and end user point of view.

7.1 Applying MPC to Real World Problems

From an end user perspective we found that MPC can both help towards enabling data access which is not possible today. However, we also found that there are some things standing in the way of doing this. Some of the things standing in the way had to do with properties of MPC technology, such as data visibility and data leakage. However, some of the things standing in the way of using data today were of a non-MPC nature. These issues were things such as internal politics, data controllers being overly cautious, incompatible data formats, different parties not knowing that others exist, etc. Hence, there is a need to understand the barriers of data sharing in the health system before deciding where and how to deploy a system using MPC technology.

We also found that if a problem is explorative in nature or full data visibility is needed, MPC is not necessarily the best solution to the problem. However, for these cases as we experienced suggestions to use it for initial hypothesis generation by using MPC to provide some initial testing that can then be used to inform whether or not to carry on with the time consuming effort of gathering the data in one place and perform data analysis in a normal way. One could also imagine that this might be applicable in more situations in which using MPC to calculate something about the data and that this might be enough to make an informed decision of the state of the data which cannot be seen due to MPC restrictions. This would of course need to be done in a way which does not leak too much information. Whether this is the case is future work.

Next, we found was that setting up an MPC collaboration requires process as well as technology. This is based on the fact that data sharing in the health domain today is hampered by poor process due to: people saying no to data access to be on the safe side (Section 5.1), unawareness that relevant data might exist in other domain than the one you operate in (Section 5.2.3), along with political reasons for not sharing data. This indicates that there is a need for a political will to change this, but also that a well-defined process for sharing data might help better use of health data across silos. Such a process is even more important when trying to bring MPC into play as this adds extra complexity. Based on the interviews (Section 5.1) and two first workshops with end users (Sections 5.2.1 and 5.2.2) we came up with such a workshop format which was tested in the third workshop (Section 5.2.3). Based on the previous, it is suggested that such a process could be widened to set up data sharing in general as well as include tools to set up MPC systems where relevant. During the workshop, we found that when discussing MPC there is a need to provide an easily understandable overview of the basic properties of MPC and what an MPC system can be used for. In that way, such a workshop could uncover potential benefits of combining data, what stands in the way of doing it, and discussion how MPC might be applied in cases where sharing data unencrypted is not an option. The workshop served well in figuring out what needs to be computed, what data sources should be utilized, and how to practically figure out how to deploy such a situation in organizational terms. For future work, we suggest to take this one step further by following up with an actual implementation of the result of the workshop.

7.2 MPC and Data Subjects

From a data subject point of view, one of the goals of the project was to explore how to create a consent interface in which privacy could be controlled. However, we quickly found that such privacy controls would not make much sense to end users. Instead we took a very explorative approach in which we set to explore: 1) what types of data the different groups generate and use in relation to their condition, 2) where this data flows, and 3) how they feel about their different types of data with the different receivers. Finally, we set out to explore how MPC would fit into or even change their privacy concerns. This was done as part of the workshops with people with dementia in the UK (Sections 5.2.1 6.2.2) and the interviews and workshops with diabetics in Denmark (Sections 5.1, 6.1.4, 6.1.3, 6.1.1, 6.1.2). We found that the groups had different views on sharing data and most surprisingly, different views on data itself.

People with Dementia While working with the people with dementia group, we discovered during the first workshop (Section 5.2.1) that it was a very special case of data subjects, or patients. The first realization was that this is a very vulnerable group which had to be dealt with carefully. Hence, a lot of the work working with this group ended up being about how to engage them in conversation about a difficult topic such as MPC. Therefore, a lot of insights into how to do this came out of the project. These are described in Escalante et al. [10]. It also turned out that discussing MPC with this group was made difficult by the fact that the participants were from the older generation. I.e., they did not have a lot of encounters with digital technology in their lives. This had made it more difficult to talk about data in the first place, let alone how it was shared. The method applied (Section 6.2.2) did, however, turn out to work well and a conversation about data, data sharing, and MPC was facilitated. For this group it meant a lot to have a sense of agency and dignity, and they wanted to know whether the technology could help them with that. So, the main focus of this group was how providing data to an MPC system could benefit them in their needs. In relation to speaking about the group as patients, they did not view themselves as patients, but rather than their condition begin a premise of their life. I.e., they saw themselves as people with dementia rather than sick people.

Diabetics While working with the diabetics, we found that they share a lot of different types of data with a lot of different organizations and people in relation to their condition (Section 6.1.1, Section 6.1.2). Examples of places they discussed different types of data about their condition were: foot specialists, eye doctors, general practitioners, and the diabetes unit at the hospital - all places that were part of their doctor mandated treatment. In addition to this, some used different apps on the phones to keep track of activities, exercise, what to eat, etc. Some also shared data with other diabetics using either groups on Facebook or by showing up to activities announced and held by the diabetes association. They were willing to share the data that was needed in order to treat their condition with the relevant parties. However, they expressed concern of employers, medical companies, and insurance companies getting a hold of their data. The group was motivated by two things when sharing their data, especially when it came to new uses such as things that would be enabled by MPC: 1) benefiting from sharing their data and 2) helping others in the same situation as themselves. When discussing what MPC might help solve, they proposed to use it to benchmark one against others with the same data. As with the people with dementia, the diabetics did also not see themselves as patients, but rather people living with diabetes as an unavoidable part of life. A final interesting observation was that since we from the beginning of the project, did not distinguish between type 1 and type 2 diabetics, both were invited to the workshop. This did create some tension, since they kept remarking

that it was two different conditions and referring to other group as "the others". They clearly did not see themselves as belonging to the same group.

Pregnant Diabetics The final group that we worked with were the pregnant diabetics which were interviewed over the phone (Section 6.1.4). In relation to the other diabetics, these were all type 1 diabetics. This group mentioned that their data sharing preferences differed depending on the context. They saw themselves as pregnant, diabetics, girlfriends, coworkers, etc. And with each context came a set of distinct data sharing preferences. This follows the theory of privacy as contextual integrity as proposed by Nissenbaum [13]. The group expressed three main reasons to share data: 1) to help research, 2) be a role model for other pregnant diabetics, and 3) that sharing data is a prerequisite for getting qualified guidance in their condition. Compared to the other diabetics, this group emphasized how much hard work there is in being a pregnant diabetic. Furthermore, their condition was no longer only about themselves but also their child. This meant that there was a higher willingness to share data if they thought it was relevant for their treatment.

Comparing the groups, we see that they are all motivated by how it can help them in their situation. Furthermore, the diabetics mentioned helping others in a similar situation as a reason for sharing data and the pregnant diabetics stated that being a role model was important. It is also interesting that both people with dementia and diabetics do not see themselves as patients, but rather as condition being a premise for their lives. Another important insight is that data subjects are very concerned with what they are getting out of contributing their data to an MPC system (or a system in general) if they are asked. This appeared within both the people with dementia (Section 6.2.2) and diabetes groups (Sections 6.1.1, 6.1.2, 6.1.3, and 6.1.4). Furthermore, the people diabetics in particular expressed a motivation to help people in similar situation, e.g., by being a role model or helping others place a device. Hence, it is important to make it clear in consent mechanism what the benefits of using a system as, as well as how contributing your data might help others. One can speculate that this differs from someone going to the emergency with a broken arm. However, this hypothesis would need to be tested, but it does have implications for designing for people with dementia and diabetics.

From the user studies activities with data subjects we learned that from the perspective of the data subjects it can be difficult to understand what MPC is and why one should be interested in it. We have also shown that the basic concepts of MPC can be explained using a video (Section 6.1.3). The question is, however, whether or not this is important or not. From the cryptographers point of view there is often an assumption that the data subjects providing data to the system should understand what is going on and why their data is, e.g., anonymous or why only your doctor and not your insurance company can the result of a computation even though they both participate in it. However, during our workshops sessions with diabetics (Sections 6.1.1 and 6.1.2) it was clear that understanding the technology was not the main concern. Instead the data subjects were concerned about concepts they could understand such as whether their data was *anonymous*. This has implications for how consent interfaces should be designed for MPC based systems. Based on terminology by Schaub et al. [14] this means that a consent interface should be designed with a multi-layered approach in which the MPC explanation is not part of the first thing a user sees, but rather something the user can unfold if interested. In this case, we would suggest that the video could be shown here. One of the workshop activities (see Section 5.2.2) during the studies indicate that MPC can be difficult to understand out of context, so we would furthermore recommend that the video and MPC explanation would be contextualized to the specific context. In addition, it should be taken into consideration that the data subjects we have studied did not see themselves as patients and should not be referred to as such. In

the communication of the consent giving experience, there should be a focus on explaining how the individual will benefit and/or explain how other people in the same situation might benefit from them contributing their data.

7.3 Trust is not equal to Trust?

In an MPC system, trust is a very important concept. Trust is often described used to describe the properties of MPC. This is done, e.g., in the case of "honest majority" protocols and a "semi-honest" setting and when describing a fundamental property of MPC systems as *eliminating the need for a trusted third party*. Furthermore, there is a tendency for cryptographers to understand trust as mathematical proof. That is, that they focus on solving the issue of trust mathematically. If it can be proven that no trust is needed from the other parties in a protocol, the problem is solved. This is, however, a too narrow view on trust. In the case of an honest majority, it specifies that as long as the majority of parties in an MPC system do not try to break the protocol to read the encrypted data, the system is secure. It does however not touch on other ways in which you have to trust the other parties. While the malicious parties might not be able to extract the data passed around in the system, they are able to pollute the result of the computation by supplying faulty data to the system. While this is not interesting or even relevant from a cryptographic protocol point of view, this limited view on trust can have implications in real world uses of MPC.

In our user studies, we found that participants had a different and often more complex view on trust. In the workshops with diabetics (Section 6.1.1 and 6.1.2) participants expressed that there were some organizations they trusted more than others. Moreover, when asked about the trust in an MPC system, they relied on institutional trust, rather than trust in the mathematical proof of an MPC system. They wanted to know who would vouch for the system, i.e., who would guarantee their anonymity (in the case that MPC is used to provide an anonymous result). In the discussion they ended up not even trusting any companies, but instead wanting the EU to guarantee the system.

Another interesting result emerged during the workshop with end users in Lancaster (Section 5.2.2) where, even though after being explained that MPC eliminates the need for a trusted third party, one of the groups introduced their own trusted third party. This third party would not receive all the data, but rather act as a mediator when negotiating what an MPC system should be able to calculate. I.e., although there might not be a need to have a party hold all the data, there might still be a need for a mediating party in which the involved parties would have to put some trust. Another type of third party emerges when deploying MPC in a real world setting. This party emerges as the system has to be implemented in software and someone has to do that. This would often be one party internal or external to the MPC system, and the other parties would have to trust that this party was able to implement the protocols correctly and that this party does not have malicious intentions.

References

[1] Akeyless. https://www.akeyless-security.com/. Accessed: 2019-12-18.

[2] Cybernetica. https://cyber.ee/. Accessed: 2019-12-18.

[3] Ibm spss statistics. https://www.ibm.com/products/spss-statistics/. Accessed: 2019-12-23.

[4] Inpher. https://www.inpher.io/. Accessed: 2019-12-18.

[5] Partisia. https://partisia.com/. Accessed: 2019-12-18.

[6] Sas software. https://www.sas.com/en_us/software/stat.html. Accessed: 2019-12-23.

[7] Sepior. https://sepior.com/. Accessed: 2019-12-18.

[8] Unbound tech. https://www.unboundtech.com. Accessed: 2019-12-18.

[9] Peter Bogetoft, Dan Lund Christensen, Ivan Damgård, Martin Geisler, Thomas Jakobsen, Mikkel Krøigaard, Janus Dam Nielsen, Jesper Buus Nielsen, Kurt Nielsen, Jakob Pagter, Michael Schwartzbach, and Tomas Toft. Secure Multiparty Computation Goes Live. In Roger Dingledine and Philippe Golle, editors, *Financial Cryptography and Data Security*, pages 325–343, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[10] Maria Luján Escalante, Monika Büscher, Emmanuel Tsekleves, Mads Schaarup Andersen, Laura Nielsen, Paris Selinas, Luke Moffat, and Jessica Robins. Ethics through Design: Medical data systems, chronically ill data subjects, and all the invisible things in between. In *Proceedings of International Association of Societies of Design Research Conference (IASDR) 2019*, 2019.

[11] O Goldreich, S Micali, and A Wigderson. How to Play ANY Mental Game. In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, STOC '87, pages 218–229, New York, NY, USA, 1987. ACM.

[12] Andrei Lapets, Nikolaj Volgushev, Azer Bestavros, Frederick Jansen, and Mayank Varia. Secure multi-party computation for analytics deployed as a lightweight web application, 2016.

[13] Helen Nissenbaum. Privacy as contextual integrity. *Wash. L. Rev.*, 79:119, 2004.

[14] Florian Schaub, Rebecca Balebako, Adam L Durity, and Lorrie Faith Cranor. A design space for effective privacy notices. In *Eleventh Symposium On Usable Privacy and Security ({SOUPS} 2015)*, pages 1–17, 2015.