

Raising legal awareness through user-centred consent visualisation

Christof Bless, Lukas Dötlinger, Michael Kaltschmid & Markus Reiter

February 19, 2021

Abstract

Data privacy scandals in the recent past have left many people around the world insecure about the implications of data sharing in everyday life. At the same time advances in machine learning and data science have given us the possibility to enhance the safety and the quality of life of many people. To collect data from individuals nowadays companies have to rely on mutual consent and trust. New regulations such as the GDPR make sure that consumers are not exploited or taken advantage of. From a legal standpoint it is necessary to inform customers to a full extent about what information is collected from them. In this paper we present a new visualisation approach to keep people informed about the activities linked to their data sharing agreements. We introduce a user-centred application with a transparent visualisation aiming to give users a better understanding of the data sharing processes in the background of their consent agreements. Finally, we show the results of a user study conducted to find out whether this visualisation leads to more legal awareness and trust.

1 Introduction

Since it has come into effect in May 2018, the General Data Protection Regulation (GDPR) [GDPR] has had a big impact on the way companies deal with personal data. The GDPR brought awareness of data ownership and protection of privacy to a new level for both companies and consumers. There is an interest to handle data generated by activities of real people in a safe and consensual way. For many companies adhering to the regulations is

not only important to prevent costly legal affairs but also to keep up a good reputation with their customers. At the same time, data owners are more and more concerned about their rights. Many value their privacy highly and want to control their digital footprint on their own.

The GDPR requires consent between a data owner and a data controller if the data controller wants to process any kind of information related to the data owner (Art. 6 (1a)). According to the GDPR, consent has to be (i) freely given, (ii) specific, (iii) informed and (iv) unambiguous (Rec. 32).

This paper focuses on establishing informed consent; That means a data owner is completely aware of extent, target and content of his data sharing activities. To achieve this we look into ways to improve transparency with the data owner through informative visualisations. We believe that visualisations help end users understand how their data is being shared, with less effort than by reading the agreement text which is the prevalent status quo.

When it comes to data visualisation to end users, the most important aspect is the simplification of complex data. Data for applications is generally stored in some sort of database system which is hard to understand for non-expert users. Data visualisation is a technique of mapping complex data to visual elements, thus making it easier to understand relations within the dataset.

This paper presents an approach to visualise data in the domain of vehicle sensor data sharing. The work is part of the CampaNeo project [**campaneo**], whose goal it is to create a system to collect and distribute sensor data generated by modern vehicles. In CampaNeo, data is requested via specific campaigns, which must be approved by the data owner. Following the GDPR, campaigns must state exactly what the purpose of their data collection is and what type of processing they plan to do on it. Ideally, companies or research organisations behind the campaigns contribute to the development of better technologies and services in the realm of mobility and transport, which in turn enhance the user's experience with the vehicle.

For example, GPS location and speed data from a big number of cars can help optimise traffic flow management, which leads to less congested roads and time savings for drivers.

Semantic knowledge graphs are a state-of-the-art solution for building versatile, explainable and machine-readable data storage solutions. Knowledge graphs are the underlying technology used in the CampaNeo project.

Since semantic data from a *triple store* mainly describes objects and their relations, it is complex to visualise. One could just display the whole database as a graph, with all objects and their relations, however, this would result in a very large visualisation which is too cluttered and therefore confusing to users.

The idea is to visualise the flow of data from a user’s car to third party companies on small to medium displays (e.g. tablet, smartphone or the car’s built-in infotainment system). The user can get an overview on what data they are sharing with institutions like governmental agencies, universities or data processing companies who collect high amounts of data with the intent to solve problems around mobility and transport. The visualisation focuses on highlighting the data streams to the user who should get information about the type of data that is shared, at what intervals it is sent out and who the receiving party is.

This paper is structured as follows: Section 1 presents an introduction to the field, while section 2 presents related work. Section 3 defines the main research questions. The methodology for deriving the first prototype can be found in section 4. ?? contains architecture details of the implementation. ?? presents the testing methodology, the results of which can be found in ?. Conclusions are made in ?.

2 Related Work

The main problem we want to solve with our work is that of fully transparent visualisation of data sharing activities. We will achieve this by building a tool that enables users to monitor and control the distribution of their data.

In recent years there have been several attempts to design applications that implement such visualisations. Raschke et al. [1] built a general dashboard to visualise data sharing activities and give consent approval and withdrawal mechanisms. The dashboard is a single page application with a vertical timeline listing the different types of actions. Among these actions are sharing a first name or a picture as well as information about location and search history. Further, the application offers information about processing context and type of the data in question.

The authors evaluated the tool with a set of tasks for participants to complete using the dashboard application. However, Raschke et al. [1] only

tested with expert users most of which were also their colleagues. The main takeaways were that data type categories need to be refined more to be understandable. Generally, even the expert users found it hard to answer questions about their data privacy based on the information available from the dashboard.

Another implementation of a consent and data privacy visualisation interface is the Consent Request (CoRe) user interface (UI) from Drozd and Kirrane [2]. The idea of the authors was to develop a UI which shows the implications of accepting a consent agreement. The hypothetical scenario would be consenting to the use of individual functionalities of a fitness tracker. For example, a user wants to have the route of a morning jog displayed on their app. To unlock the functionality, one has to accept some data processing by the data controller, which is the manufacturer of the tracker. The CoRe UI [2] will then display a graph that shows what data is sent out, where it will be stored, the type of processing that is done on it and which third party companies it will be shared with.

To validate their design choices, Drozd and Kirrane evaluated two slightly different CoRe UI prototypes by giving tasks to participants from different age groups and recording their actions. The first prototype that was tested was a bit more elaborate and had more features than the second one which was a simplified version. For the first test 27 participants were asked to perform the specified tasks. 74% of participants seemed to be “very confused” and more than half of them found the UI “too complex” and “hard to use”. The second, simplified UI was accepted better with an even larger test group of 74 people. Still many would describe the layout as being “confusing”, “annoying” and “complex”. The majority of participants claimed to be unsatisfied or neutral with the application.

The tolerance for cognitive overload through too much information and display of complex relations is low for most people. Especially when dealing with legal conditions of data privacy. Consequently, a good start for any attempt to create a transparent visualisation of data sharing processes is to simplify the user interface to only include the most essential components, which will be determined in Section 4.

3 Research Question

The CampaNeo project is highly dependent on requesting and receiving informed consent for sensor data sharing. The more people agree to send usage data from their cars, the more value the statistical analysis will generate. We formulate two hypotheses which are the basic assumptions of this work:

- People are more willing to share their data if they are fully informed on what exactly they are sharing, when they are sharing it and with whom exactly they are sharing it.
- Data visualisations improve comprehension of consent.

It is debatable whether current consent gathering methods really make it absolutely clear to data subjects what happens in the background after they gave their consent. According to [3], people tend to agree to most consent requests they are confronted with. Reading through all the agreement specifications is time-consuming. Such documents are often written in a complex language typical for legal documents. Due to that, most people who give their consent to data sharing agreements do so without understanding many details of the contract. Bechman [4] defines this as a “culture of blind consent”. To conclude, in most cases having one’s consent, even informed, is not equivalent to having awareness.

To change this, we build an application to enable data owners to give informed consent and gain legal awareness in the process. This should be achieved by making the data flow completely transparent through a visualisation.

The research questions that are addressed in the design process are:

1. What aspects of the data should be visualised?
2. How can the data be visualised in order to improve comprehension?

4 Prototype