

Towards a better modelling of consent

Data sharing is not a new phenomenon. With the current technological advancements it is happening everywhere and at any time. But do we know which data is actually shared, with whom, where and for which purposes? For many years already, consent for data sharing has been addressed by both users and companies. With the enforcement of the GDPR¹ directive in 2018, however, user rights and data privacy have shifted into the focus of attention. In particular article 14² of the GDPR directive which introduces the notion of informed user consent and penalties for non-compliance. Fines may add up to 10 million euros or 2% of the company's worldwide annual revenues (Article 83³). Furthermore, the issue of informed user consent is becoming more and more prominent, especially in the current year of 2020, which confronted us with a pandemic leading to an adverse situation and many problems. More often than ever people are asked to share personal information with authorities (e.g. location tracking, health information), which is then analysed and shared with third parties. All this leads to a major shift in all sectors that depend on data sharing.

Some of the major challenges many companies encounter during this adjustment period to GDPR are representing and managing consent. Representing consent can be looked at from many perspectives. From a technological point of view, consent is an agreement between two enti-

ties, namely a user and a company. But is this enough? The answer is no. Research has shown that most of the users are unaware of what it means to consent and the implications of giving it. Of course, there are users who are aware of the problem, but external and internal factors may affect their judgement. With the boom of social media more and more people feel pressured to fit in, thus a culture of a "blind consent"⁴ has developed. Privacy policies are disregarded and almost never read due to being too long, too complicated and written in a legal jargon, which is not aimed at the every-day user. All in all, GDPR calls for transparency at each level of the data sharing process: from the type of data itself and processing applied to it all the way to the metadata and algorithms associated with it. Consent can be requested via GDPR compliance forms but representing it in both machine-readable and human-readable form is a challenge. How do we make sure that users understand what actually happens to their data? Who has access to the data and who has not?

Research is on the rise and solutions using machine learning and artificial intelligence are in development. For example, such work is currently being done in the FFG CampaNeo⁵ and Horizon 2020 smashHit⁶ projects partnered by the University of Innsbruck.

CampaNeo's main goal is to set up a platform for secure vehicle sensor data sharing between multiple entities. The platform is to provide sensor data analysis and future predictions generated with machine learning models. The sensor data are collected in real-time from a dedicated fleet of cars of Volkswagen that are equipped specifically for the purpose. The smashHit project, funded under H2020-EU.2.1.1 builds further upon the solution of the CampaNeo and provides users and industry with a secure data sharing platform. smashHit addresses consent, its representation and management on a deeper level and has planned to have a dedicated tool for automatic contracting.

Both CampaNeo and smashHit address consent, but how exactly will a solution be developed? In the past decade, semantic technologies, namely ontologies and knowledge graphs, started to gain broad real-life adoption. Due to their ability to represent knowledge in human-readable and machine-readable format, interoperability, faster and easier knowledge discovery, big companies such as Google have built knowledge graphs and integrated them in their systems. This is also the case with the CampaNeo and smashHit projects, whose data models are represented as knowledge graph. In both projects, consent is one of the main priorities and will be represented with a carefully crafted ontology. Several ontologies for consent, such as the CDMM Consent On-

1 gdpr-info.eu

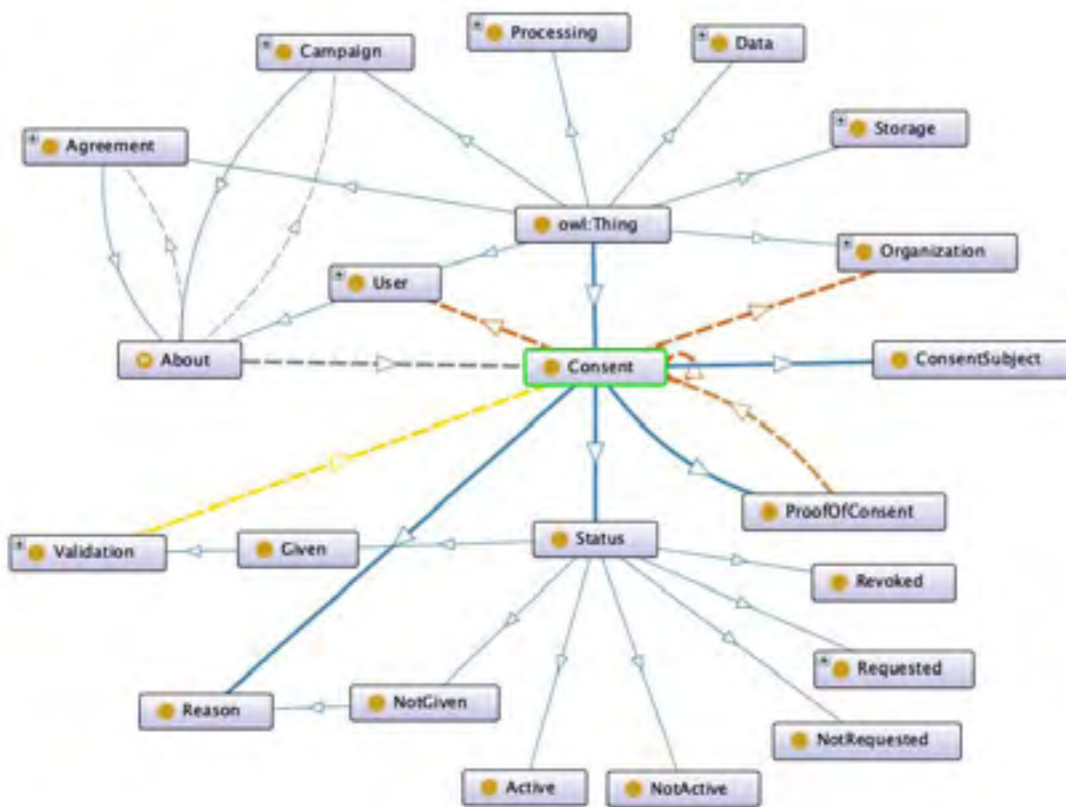
2 gdpr-info.eu/art-14-gdpr/

3 gdpr-info.eu/art-83-gdpr/

4 A. Bechman, Non-informed Consent Cultures: Privacy Policies and App Contracts on Facebook, *Journal of Media Business Studies* 11 (1): 21-38, 2014

5 projekte.ffg.at/projekt/33146680

6 www.smashhit.eu



tology⁷ and GConsent⁸ have already been developed in the past and even though they have detailed representation of consent according to GDPR, there are limitations. In most cases, the ontologies represent the notion of consent but do not provide information about which data consent is required for, for which purposes, i.e. handle data granularity in a restricted way, especially for the sensor data, and are not able to handle consent state changes. With smashHit and CampaNeo we aim to overcome these limitations by reusing several ontologies from different areas (finance, business, legal, technology, engineering, mobility) and to build a complete solution that focuses on consent, users, data, processing. Third-party organizations thus provide full transparency of the data sharing process.

The diversity of partners in the smashHit project allows for the unique opportunity of testing the developed solution in the

7 www.w3.org/community/dpvcg/wiki/CDMM_Consent_Ontology

8 opscience.adaptcentre.ie/ontologies/GConsent/docs/ontology

real-world. Infotripla⁹ and Forum Virum Helsinki¹⁰, two of the main industry partners in smashHit, have provided several use cases based on real-world challenges that they have encountered since the acceptance of GDPR in the smart city domain. In addition, Volkswagen¹¹, Atos¹², LexisNexis¹³ and the X/OPEN Company¹⁴ provide their expertise in the areas of mobility, digital transformations, and insurance.

In conclusion, the expansion of the Internet of Things has allowed connecting everyday devices to a network that can be accessed at any time and has defined data as a new currency. Informed user consent is not only expected but also obligatory for GDPR compliance and the user acceptance of the technology. Institutions should focus on raising awareness as well as on informing about possible

9 www.infotripla.fi

10 forumvirium.fi/en/frontpage-2/

11 www.volkswagenag.com/

12 atos.net/en/

13 www.lexisnexis.com/en-us/gateway.page

14 www.opengroup.org/unix@-systems

implication and user rights. Here the semantic technology can provide better modelling solutions.



Anna Fensel is Associate Professor at Semantic Technology Institute Innsbruck, Department of Computer Science, University of Innsbruck, Austria. She has her habilitation and her doctoral degree in Computer Science from the University of Innsbruck, and a Mathematics and Computer Science degree from Novosibirsk State University, Russia.



Anelia Kurteva is working at Semantic Technology Institute (STI) Innsbruck. Her focus is knowledge representation and visualisation in the area of data governance and privacy for smart cities.