Data sharing in smashHit: Making consent and contracts interpretable with knowledge graphs

Anelia Kurteva, Amar Tauqeer and Anna Fensel

The General Data Protection Regulation (GDPR)[1] has caused a technological shift in how data processing and sharing are performed. Consent, as a legal basis for data processing, has become a focus for many. However, it is only one of the six legal bases of GDPR. Contracts or contractual obligations are also a way of lawfully requesting data for various purposes. Although severe fines have already been imposed, many companies still struggle to comply with GDPR and when they do compliance is rarely done at scale. This can be due to the lack of resources and technology to simplify the compliance process and legacy systems built before the GDPR. Another major challenge is the lack of standard data models for both consent and contracts, which are in both machine-readable and human-readable formats. In smart cities, for example, where Internet of Things (IoT) data-centric solutions are rapidly being developed, establishing a common understanding of legal requirements for data processing such as consent and contracts is key.

The Semantic Web can provide a solution to these problems, and it is called knowledge graphs. Knowledge graphs, which Fensel et al. [1] define as *"very large semantic nets that integrate various and heterogeneous information sources to represent knowledge about a certain domain of discourse"*, have been around for a while, but their applications and benefits for smart cities, and specifically for managing informed consent and contracts are yet to be explored. Due to their ability to transform data into information and information into knowledge by adding context via relationships between entities, knowledge graphs can provide the level of traceability, transparency and interoperability that is needed for the implementation of consent through its entire life-cycle [2] (i.e., consent request, comprehension by machines and humans, consent decision, the use of it and its withdraw). In a similar manner, knowledge graphs can be used for modelling contracts, which in comparison to consent are more complex and can go through several negotiation phases.

Enabling data interoperability in smart cities is also one of the main goals of the FFG CampaNeo[2] and the Horizon 2020 smashHit[3] projects. 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 builds further upon the solution of the CampaNeo and provides customers and businesses with a secure data sharing platform. smashHit addresses consent and contracts and their semantic representation and management on a deeper level. In the smashHit project, consent and contracts are main priorities and are/will be represented with a carefully crafted ontology (**Fig.1** and **Fig. 2**). Our survey [2] on semantic models for consent has shown that there is a variety

[1] https://eur-lex.europa.eu/eli/reg/2016/679/oj
[2] https://projekte.ffg.at/projekt/3314668
[3] https://www.smashhit.eu

of ontologies that can be reused and adapted for our use cases. Several ontologies for consent, such as the CDMM[4] Consent Ontology and GConsent[5] have already been developed and even though they have a detailed representation of consent according to GDPR, some limitations still exist. In most cases, the ontologies represent the notion of consent, but do not provide information about what data is consent required for, for what 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 build a holistic and extensible solution that focuses on consent, users, data, processing, third-party organisations, providing full transparency of the data sharing process. However, although consent is modelled in detail, contracts are rarely represented by the same semantic model. This is what makes our smashHitCore ontology stand out.

smashHitCore goes beyond consent and addresses contracts, which are another GDPR legal basis for data processing. In this way, smashHitCore provides a flexible yet meaningful model for data sharing in smart cities under different circumstances. In addition, the efforts on the smashHit project have resulted in a data protection by design tool for automated GDPR compliance verification based on semantically modelled informed consent (see [3]). With our solution for compliance in [3], we have shown the successful application and use of legal knowledge graphs in complex scenarios such as GDPR compliance and the benefits that come with that use.

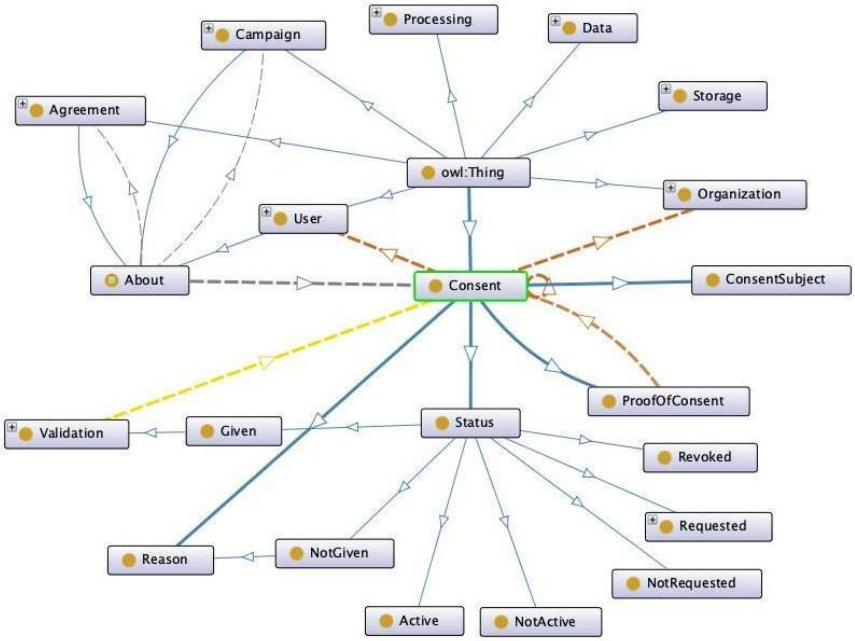

Fig. 1 Overview of the smashHitCore ontology for secure and controlled sharing of personal and industrial data.

[4] https://www.w3.org/community/dpvcg/wiki/CDMM_Consent_Ontology
[5] https://openscience.adaptcentre.ie/ontologies/GConsent/docs/ontology

Providing a service for automated contracting in compliance with GDPR is another goal of smashHit. The main challenge is to bind GDPR rights with businesses, specifically in the form of digital semantic contracts. This creates an opportunity for small and medium enterprises (SMEs) to manage personal data in contracting services at scale. For this to happen, organisations have to adopt security and privacy measures as outlined in GDPR. Art. 28 (3). By following these requirements and actively collaborating with both industry and legal experts, we are building a solution for contracting, which supports the detection of contract breaches, contracts traceability and data interoperability.

For the semantic representation of contracts, we have reused the Financial Industry Business Ontology (FIBO)[6] [4]. We have reused two major classes: Contract and Mutual Contractual Agreement, which are generic enough to model diverse use cases. Similarly to consent, contracts can have multiple states, which we have modelled as well as shown on Fig 2.

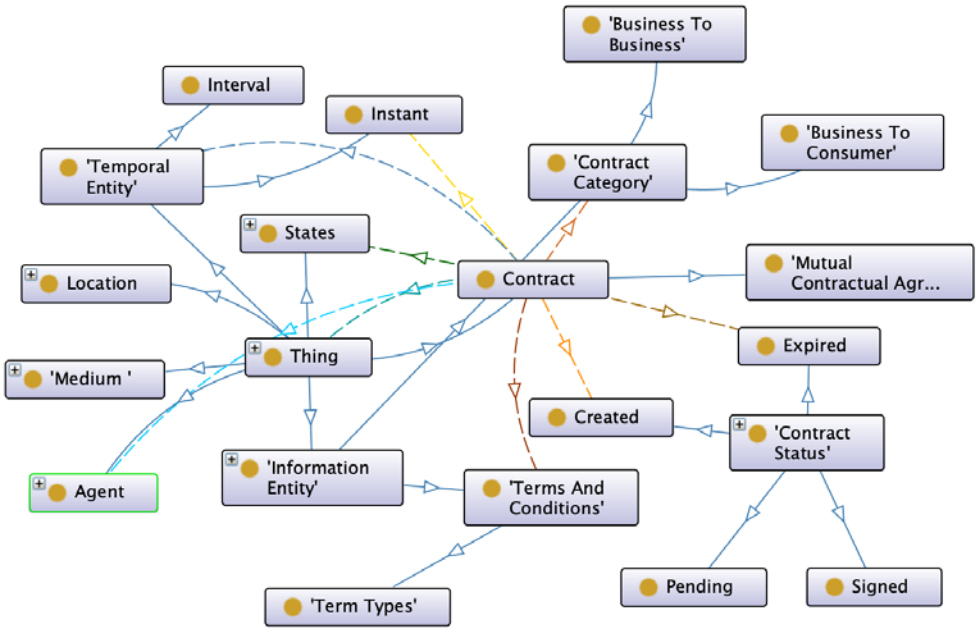

Fig. 2 Overview of the smashHitCore semantic model for contracts

The diversity of partners in the smashHit project allows for the unique opportunity of testing the developed solution in the real-world. Infotripla[7] and Forum Virum Helsinki[8], 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

[6] https://spec.edmcouncil.org/fibo/
[7] https://www.infotripla.fi
[8] https://forumvirium.fi/en/frontpage-2/

domain. In addition, Volkswagen[9], Atos[10], LexisNexis[11] and the X/OPEN Company[12] provide their expertise in the areas of mobility, digital transformations, and insurance.

In conclusion, the expansion of the IoT has allowed connecting everyday devices to a network that could be accessed at any time and has defined data as a new currency. Informed consent is not only expected but also obligatory for GDPR compliance and for individuals' acceptance of the technology. Institutions should focus on raising the awareness regarding it, its possible implications, legal rights, and how semantic technology can deliver better modelling solutions.

References

1. Fensel D. et al. (2020) Introduction: What Is a Knowledge Graph? In: Knowledge Graphs. Springer, Cham. https://doi.org/10.1007/978-3-030-37439-6_1
2. Kurteva, A., Chhetri, T. R., Pandit, H. J., & Fensel, A. (2021). Consent through the lens of semantics: State of the art survey and best practices. *Semantic Web*, (Preprint), 1-27.
3. Chhetri, T.R., Kurteva, A., DeLong, R.J., Hilscher, R., Korte, K., Fensel, A. Data Protection by Design Tool for Automated GDPR Compliance Verification Based on Semantically Modeled Informed Consent. *Sensors*. 2022; 22(7):2763. https://doi.org/10.3390/s22072763
4. Petrova, G., Tuzovsky, A.F., Aksenova, N.V.: Application of the financial industry business ontology (fibo) for development of a financial organization ontology. In: Journal of Physics: Conference Series. vol. 803 (2017)

About authors:

Anelia Kurteva and **Amar Tauqeer** are PhD students at Semantic Technology Institute (STI) Innsbruck, Department of Computer Science, University of Innsbruck, Austria.

Anna Fensel (expected to be a presenter at the workshop) is Associate Professor at Wageningen University, the Netherlands, and at Semantic Technology Institute (STI) Innsbruck, Department of Computer Science, University of Innsbruck, Austria.

All authors are working on Horizon 2020 smashHit project (grant agreement 871477).

[9] https://www.volkswagenag.com/
[10] https://atos.net/en/
[11] https://www.lexisnexis.com/en-us/gateway.page
[12] https://www.opengroup.org/unix-systems