Software Impacts 21 (2024) 100653

Contents lists available at ScienceDirect

Software Impacts

journal homepage: www.journals.elsevier.com/software-impacts

Original software publication

GDPR Data Sharing Contract Management and Compliance Verification Tool

Amar Tauqeer [a,b,*], Anna Fensel [b]

[a] Semantic Technology Institute (STI), Department of Computer Science, Universität Innsbruck, 6020 Innsbruck, Austria
[b] Artificial Intelligence Chair Group, Wageningen University & Research, The Netherlands

ARTICLE INFO

Keywords:
Data sharing
Contracts
Research data
Semantic technologies
GDPR compliance
Smart cities
Insurance

ABSTRACT

General Data Protection Regulation (GDPR) is compulsory for processing personally identifiable data (PID) across Europe after 2018. Specifically when sharing research data, we cannot process PID without a legal basis defined by GDPR. To facilitate this, we present a scalable and interoperable automated Contract Compliance Verification (CCV) tool that enables GDPR-compliant contract management and data sharing. With the implementation of two scenarios in smart cities and insurance domains, we show how CCV is helpful to process PID and make data collection and integration (*e.g.*, from crowdsourcing) easier.

Code metadata

Current code version	v1
Permanent link to code/repository used for this code version	https://github.com/SoftwareImpacts/SIMPAC-2024-95
Permanent link to Reproducible Capsule	
Legal Code License	http://creativecommons.org/licenses/by/4.0/
Code versioning system used	Git
Software code languages, tools, and services used	Docker, Python, React, Next, flask, RDF, GraphDB, SPARQL.
Compilation requirements, operating environments & dependencies	Python 3.9+, SQLite, Ubuntu 22.04, etc.
If available Link to developer documentation/manual	
Support email for questions	amar.tauqeer@wur.nl

1. Introduction

Data sharing has an enormous impact on every type of industry in the modern, digitized world. An enchanting scenario is the sharing of research data among researchers in the digital universe where research data are frequently not shared [1]. Institutes are evaluating their privileges, roles, and duties for overseeing and for exploiting research data from their researcher [2]. In domains, such as healthcare, smart cities, insurance, and autonomous vehicles, numerous opportunities have been enabled by data sharing [3,4]. Another scenario in which data sharing played a crucial role is the case of COVID-19 contract tracing application, which assisted in restricting the virus's spread and thereby saving lives [5]. The smart city is another example, where data sharing is compulsory to enable smart city components, such as smart mobility and a smart environment [6]. However, despite how crucial data sharing is in the modern digital age, privacy, and data misuse have become big issues.

This growth in privacy concerns has resulted in the development of various privacy protections, particularly in legislation. The General Data Protection Regulation (GDPR) [7,8], which was implemented by the European Union (EU) on May 25, 2018, is one of the prominent examples of privacy law. The GDPR [7,8] outlines the six legal bases (*i.e.*, consent, contract, legal obligation, vital interest, public task, or legitimate interest (GDPR (Art. 6))) for processing personally identifiable data (PID) and is applicable to all parties handling data belonging to EU citizens. To handle any PID, prior consent from the data subject (data owner) must be obtained. Sharing research data among researchers is also an important scenario where consent is not enough and a contract is necessary. A contract is also necessary for further scenarios like the selling of data between organizations, where consent is not enough [9]. As this tool is mainly focused on data sharing through contracts that comply with GDPR so we continue our discussion with this legal basis.

For lawful data processing, consent is not the only legal basis. Another important GDPR legal basis is "Contractual performance",

* Corresponding author at: Artificial Intelligence Chair Group, Wageningen University & Research, The Netherlands.
E-mail addresses: amar.tauqeer@wur.nl (A. Tauqeer), anna.fensel@wur.nl (A. Fensel).

https://doi.org/10.1016/j.simpa.2024.100653
Received 17 April 2024; Received in revised form 24 April 2024; Accepted 3 May 2024

which is defined as "processing is necessary for the performance of a contract to which the data subject (DS) is party or in order to take steps at the request of the DS prior to entering into a contract" GDPR (Art. 6(b), Rec. 44). Following this legal basis, GDPR also has introduced other additional requirements such as the implementation of the necessary technical and organizational measures (TOMs) or data protection by default (Rec. 78) and duties of the Data Controller (DC) or the Data Processor (DP) for lawful data processing. For the fulfillment of these requirements, the following significant challenges are posed by GDPR: (1) automatic compliance verification (*i.e.*, contract compliance verification in our case), (2) the implementation of GDPR data protection default principles, and (3) the translation of the legal requirements such as purpose limitation and the minimization of storage into code. Along with GDPR, the other key challenges are scalability and interoperability [10].

As mentioned in [11], the EU companies are only 47% "fully" or "very" compliant, there is a need for a technical solution that overcomes the challenges raised by GDPR. This solution not only provides safeguards for personal preserves but also unlocks the data-sharing benefits in the modern digitized world, more specifically in digital contracting services and for sharing research data among researchers. In addition, the tool also makes data collection and integration (*e.g.*, from crowdsourcing) easier as the tool ensures that (research) data sharing complies with GDPR.

Here, we present a tool that aids data sharing contract management functions such as semantic contract creation, contract audit, and automated contract compliance verification checks to enable GDPR-compliant data sharing and processing in research (sharing research data) and that is specifically adapted for and tested in domains of smart cities and insurance. Our tool complies with the following GDPR [7,8] requirements: GDPR (Art. 5(1)(a)) "lawfulness, fairness, and transparency", GDPR (Art. 5(1)(b)) "purpose limitation", GDPR (Art. 5(1)(c)) "data minimization", and GDPR (Art. 5(1)(d), Art. 5(1)(f), Art. 5(2)) "accuracy, integrity, confidentiality, and accountability". Further, our tool is implemented with the principles of data protection by design as it is a key requirement of GDPR. Our tool brings the features like auditability and interoperability with the use of contractual data in knowledge graphs (KGs) [12,13] for the semantic contract representation.

The tool has been successfully tested with two real-world use cases, namely, Smart City Services (UC1) [14] and Insurance Services (UC2) [15] in the smashHit[1] project, for data sharing in the smart city and insurance domains. The smashHit project's goal was to develop a scalable, reliable, and secure system for managing consent and contracts for data sharing that complied with GDPR in the connected automobile and smart city domains. In order to enable GDPR-compliant data sharing through contracts, we deduce GDPR-compliant requirements from both use cases. In the smashHit project, our tool is referred to as the automatic contracting tool (ACT) [16,17]. Since our tool enables data sharing and processing along with GDPR, therefore it is helpful for any type of industry that involve in data processing and sharing, specifically in research for sharing research data. Specifically, data updates, in the case of a contract breach, the tool sends notifications automatically to the contractors who are part of that particular contract.

2. Software description

To enable scalability, our tool uses the microservices architectural pattern. Docker,[2] a containerization technology used for tool portability and agility. Python[3] is used for developing this tool. Section 2.1 details the software architecture, while Section 2.2 describes the software's core functionalities. Our paper [9] and smashHit white paper [16] provide further detail about the tool.

2.1. Software architecture

A high-level overview of the software architecture of the data sharing contract management and compliance verification tool along with GDPR is depicted in Fig. 1, while the interaction of the tool's components with the external entities such as the DC, DP, and DS is shown in Fig. 2. The tool provides the REST (REpresentational State Transfer) API (application programming interface) endpoints, which allow other applications and software to interact with the data sharing contract management module (*i.e.*, backend core and compliance module). The contract management module further interacts with other microservices such as the compliance scheduler and GraphDB[4] (the graph database for storing data). The contractual information represented in KGs is stored in GraphDB. To perform Create, Read, Update, and Delete (CRUD) operations on contractual data stored in KGs, the SPARQL[5] (SPARQL Protocol and RDF (Resource Description Framework) query language) is used. Moreover, to protect PID, it must be encrypted before saving it into GraphDB, and on requesting it should be decrypted. For this purpose, the tool implements a hybrid layered encryption scheme using Rivest–Shamir–Adleman (RSA) [18] and Advanced Encryption Standard (AES) [19]. This layer makes the query of the encrypted contractual information without showing any PID from GraphDB.

2.2. Software functionalities

The main functionalities of our tool are as follows.

2.2.1. User authentication

To access the tool, users must first sign-up and be authenticated via the contract REST API endpoint.[6] A JSON Web Token[7] (JWT) is required to access any contract REST API endpoint,[8] which can be generated on successful authentication.

2.2.2. Data protection by design

The tool is built on the principle of data protection by design and facilitated features like securing queries while making over encrypted contractual information stored in the GraphDB and encryption scheme using the RSA and the AES algorithms. In addition, the contract REST API endpoints are secured with a JWT token.

2.2.3. Data sharing contract management

After successful authentication, end users (*i.e.*, contractors, organizations, and software applications) can perform data-sharing contract management functionalities such as creating, updating, deleting, and reading semantic contracts via the contract REST API endpoints. As contract lifecycle management comprises many stages [9], the end users have to manage the contractual information in each stage. The tool supports the management of contractual information, such as the contract's basic information (*i.e.*, contract category, contract type, contract purpose, and contract dates (effective date, end date)) contractual terms, contractual clauses (*i.e.*, contractual obligations), and contractors' information (*e.g.*, name, address, signatures).

2.2.4. Contract compliance verification

The tool performs automated contract compliance verification checks on semantic contracts based on the CCV (Contract Compliance Verification) scenarios discussed in our paper [9]. An example of the CCV scenario (Business-to-Consumer (B2C) data sharing contract) is illustrated in Section 3. The notifications about contract compliance verification breaches/violations are sent to contractors automatically by the tool through email (see the result of a contract breach/violation message in Fig. 4).

[1] https://ec.europa.eu/isa2/solutions/european-union-public-licence-eupl_en/
[2] https://www.docker.com/
[3] https://www.python.org/
[4] https://www.ontotext.com/products/graphdb/
[5] https://www.w3.org/TR/rdf-sparql-query/
[6] https://actool.contract.sti2.at/swagger-ui/
[7] https://jwt.io/introduction
[8] https://actool.contract.sti2.at/swagger-ui/

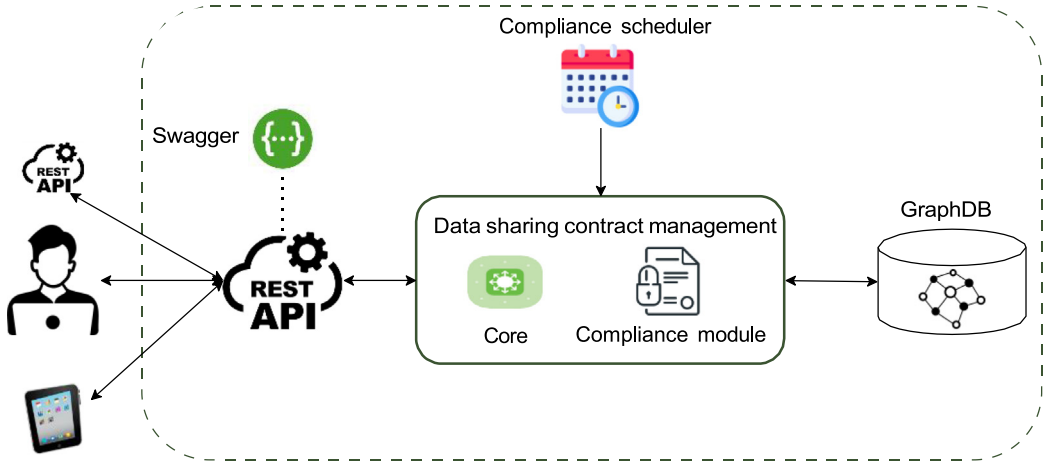

Fig. 1. An overview of the software architecture of data sharing contracts management and compliance verification.

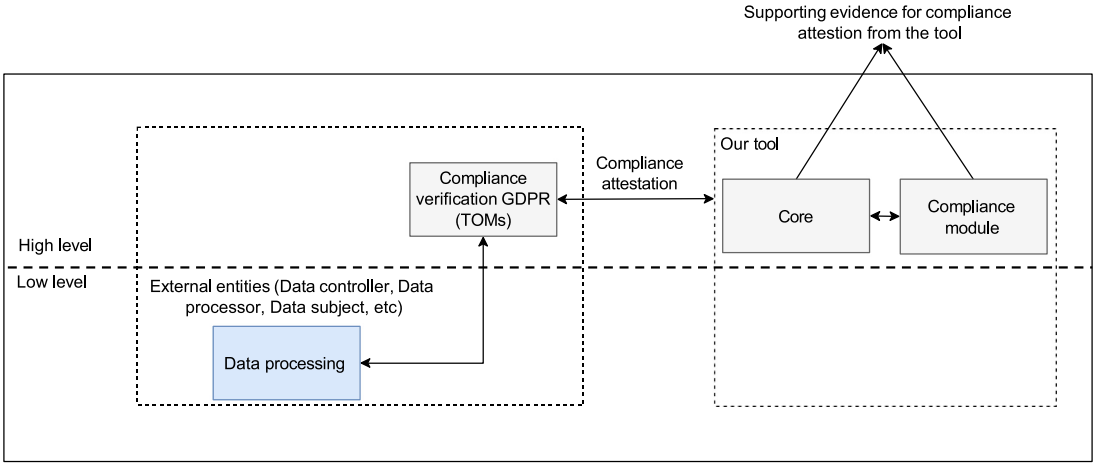

Fig. 2. Interaction between external entities and our tool's component for contract compliance verification.

2.2.5. Interoperability

The interoperability of the tool is maintained by KGs [16], which gives a persistent and homogeneous representation of semantic contracts. Furthermore, the tool's KGs and ontology [20] were created in consultation with the smashHit project's legal team and industrial partners.

2.2.6. Triggering notification

The tool performs automated and manual contract compliance verification to detect contract breaches/violations. Whenever a contract breach is detected, the tool sends notifications automatically to the contractors (involved in that particular contract) about the contract violations/breaches via email.

2.2.7. CRUD operations

Our tool provides the facility to perform CRUD operations on data-sharing contracts. Both partial audit and full audit are available in the tool. Since the contract creation process is divided into different parts, therefore, CRUD operations can be performed on each part of a contract such as contractual terms, contractual obligations, and contractors' data.

3. Illustrative example

In the tool, users must create a data-sharing contract to perform contract management and GDPR contract compliance verification. A contract can be made via the contract REST API endpoint (*i.e.*, create see Fig. 3).

To illustrate, let us take a B2C data-sharing contract between Scott (*i.e.*, a DS (*e.g.*, a person who wants to share research data)) and LexisNexis (*i.e.*, a DC who wants to process research data). The duration of this contract is nine months from 07-09-2022 to 07-07-2023. Both contractual parties have agreed upon the contractual terms and conditions (including clauses) and signed the contract. The input passes to our tool via API in JSON form for creating this data sharing contract, which is shown in Listing. 1. In addition, Fig. 5 shows the input through the CURL command and the response of the data-sharing contract creation. While a KGs representation of the data-sharing contract (B2C) is shown in Fig. 6. Within the provided dates, the tool shows there is no violation. Let us change the current date to 07-08-2023 and execute the contract compliance endpoint to see the violation result which is presented in Fig. 4. The violation message shows that the contract end date is passed, so the DC cannot process data anymore. The tool sends this violation message to the contractors automatically by email.

```
{
  "consentId": "",
  "consideration": "data sharing between
    Scott and

     LexisNexis",
  "contractCategory": "
    categoryBusinessToConsumer",
  "contractStatus": "statusCreated",
```

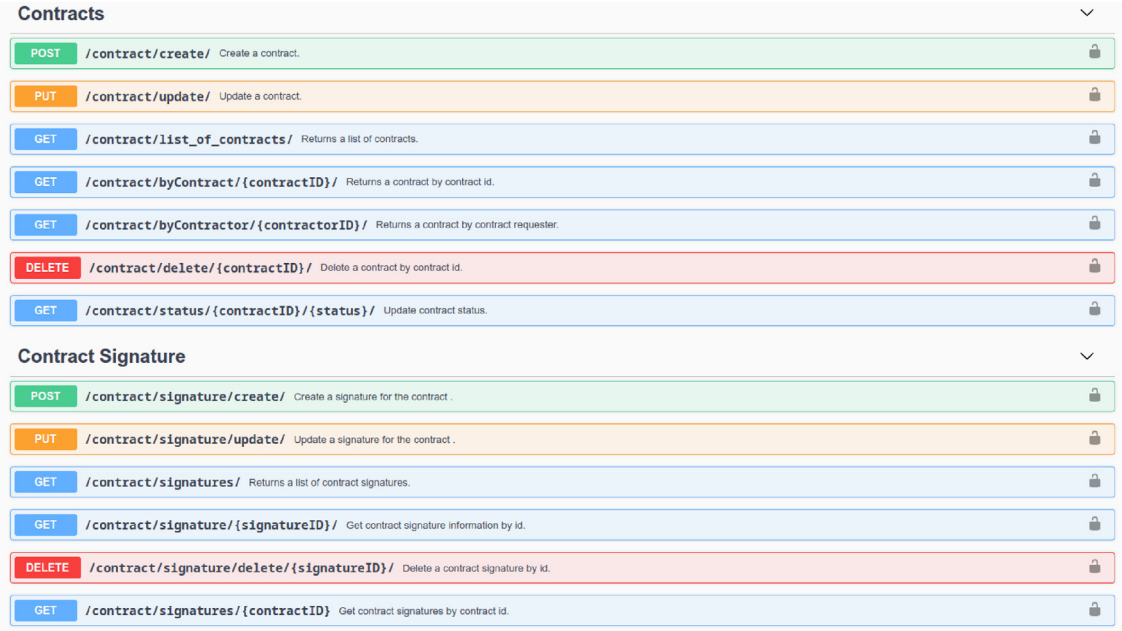

Fig. 3. A screenshot of contract REST API endpoints (only for contract and contractor signatures).

```
  "contractType": "Written",
  "effectiveDate": "2022-09-07
    10:38:07.617000+00:00",
  "endDate": "2021-07-07
    10:38:07.617000+00:00",
  "executionDate": "2022-09-07
    10:38:07.617000+00:00",
  "identifiers": {
    "contractors": [
      "c_356d371c-2e97-11ed-be7d-3
    f8589292a29",
      "c_6a094420-2e97-11ed-be7d-3
    f8589292a29"
    ],
    "signatures": [
      "sig_0d930c10-2e99-11ed-be7d-3
    f8589292a29",
      "sig_1c5e38e6-2e99-11ed-be7d-3
    f8589292a29"
    ],
    "obligations": [
      "ob_9f218204-2ed2-11ed-be7d-3
    f8589292a29"
    ]
  },
  "medium": "online",
  "purpose": "data sharing between Scott and
     LexisNexis",
  "value": "1000"
}
```

Listing 1: Data sharing contract request JSON schema (taking input as a JSON).

In addition, more elaboration on use cases in the insurance and the smart cities domain where our tool is tested can be found in the smashHit white papers [21,22].

4. Impact

The tool is a conceptualized and implemented solution as a trusted, secure, and integrating privacy-by-design reference framework to simplify the contract process, as well as to enable contract tracing and sharing among multiple platforms, specifically sharing research data. To develop sector-specific and cross-sectoral services, it ensures and safeguards (through joint security and privacy-preserving mechanisms) the sharing of data streams from both personal and industrial platforms. Our tool highlights the limitations such as scalability, interoperability, contract management, contract compliance verification, and the implementation of TOMs, thereby, enabling data-sharing, which plays a vital role in preserving personal privacy. Further, it offers features like automatic deduction of contract violations along with GDPR, and the management of contract lifecycle (the ability to perform CRUD operations on the data sharing contracts). Since the solution is built with semantic technology and has semantic models of modern digital contracts, the information is available in a machine-readable format and can be used in various contracting tools or further improved to be suitable for other domains, specifically in research (for sharing research data).

The study has been conducted and the solution has been implemented with two real-world use cases: UC1 [14] and UC2 [15] in the smashHit project domains, such as insurance and smart cities: this is documented in our previous works [9,10,20]. UC1 [14] focuses on insurance services that rely on vehicle sensor data sharing. To share vehicle sensor data (containing PID) with third parties in a manner compliant with GDPR is necessary in the modern digital world. Our tool helps in the management and sharing of this vehicle sensor data (that complies with GDPR) with third parties by providing consent and contract-based data sharing along with GDPR. In UC2 [15], Helsinki follows MyData [23] principles where residents must be able to manage PID that Helsinki collects (*e.g.*, authorize its use for different purposes and services, be able to restrict and even deny access to it). PID (*e.g.*, vehicle sensor data such as fuel, speed, location) is essential for the current traffic planning and management digital services [20]. The tool solves this real-world problem by managing the PID of Helsinki residents in a GDPR-compliant manner using consent and contract-based data sharing that complies with GDPR, with the aim that they have more control over data.

The solution has been implemented with two real-world use cases: UC1 [14] and UC2 [15]. However, the solution is not limited to these use cases but also with some customization in the implementation of the tool is feasible for other domains *e.g.*, health for making data collection and integration (*e.g.*, from crowdsourcing) easier.

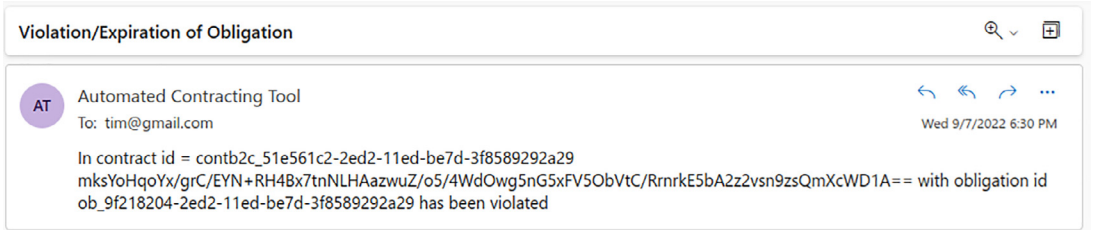

Fig. 4. A B2C contract violation result.

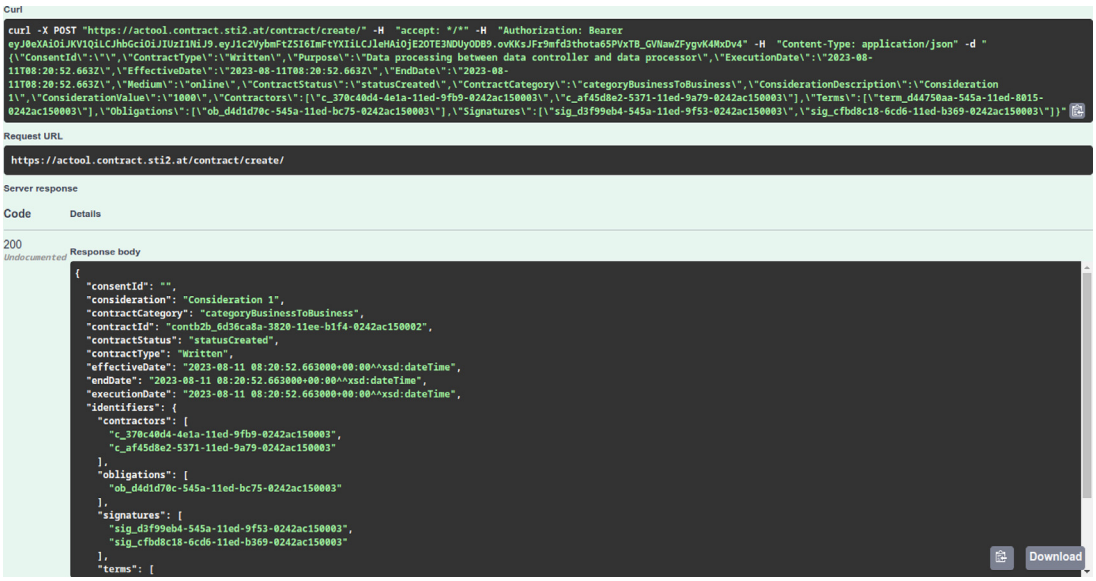

Fig. 5. A data-sharing contract creation via API and its response.

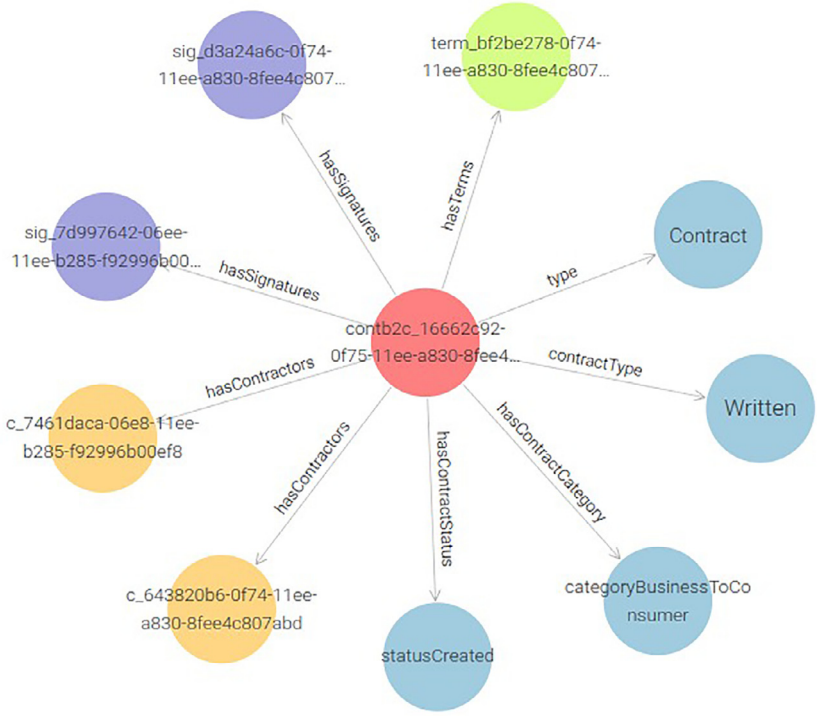

Fig. 6. KG representation of a B2C contract that is stored in GraphDB in an encrypted format.

5. Conclusion

In this work, we have presented a scalable and interoperable tool for managing data sharing contracts and automated CCV ensuring compliance with GDPR. The industrial partners and the legal experts were involved in the development of this tool, where the tool was tested against real-world use case scenarios. Data processing and data sharing, particularly in industrial and academic in research, can benefit from this tool, as the tool enables GDPR-compliant data sharing. Our tool provides more control over data to the individuals by helping them to be aware of whether the data processing is lawful as per the data-sharing contract (*e.g.*, in research for sharing research data). Further, the tool makes data collection and integration (*e.g.*, from crowdsourcing) easier as it ensures compliance with GDPR. The source code for the tool is available on GitHub[9] under MIT license. The following are included in the future development of the software: (1) comprehensive documentation of the tool, (2) an elaborated graphical interface, (3) verifications of the data sharing contracts through digital signatures, (4) improvement in digital assets licensing through DALICC[10]/Licence Clearance Tool (LCT),[11] (5) a graph-based data validation using Shapes And Constraints Language (SHACL), and (6) optimization of the tool performance.

Publications

The scientific publications enabled by this software:

- A. Tauqeer, A. Kurteva, T. R. Chhetri, A. Ahmeti, A. Fensel, Automated gdpr contract compliance verification using knowledge graphs, Information 13 (10) (2022). doi:10.3390/info13100447.

Funding

This research is supported by the smashHit EU project funded under Horizon 2020 (grant number 871477).

CRediT authorship contribution statement

Amar Tauqeer: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Anna Fensel:** Writing – review & editing, Writing – Supervision, Project administration, Methodology, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We express our gratitude to Friederike Knoke and Samuel Iheanyi Nwankwo from the University of Hannover for the legal analysis of the contracts in our use cases. We also appreciate our colleagues Tek Raj Chhetri, Albin Ahmeti, Robert David, and Geni Bushati for their assistance in developing the tool. Finally, we thank our industry collaborators LexisNexis Risk Solutions, Volkswagen AG, Infotripla and Forum Virium Helsinki for supporting the use cases for our work.

[9] https://github.com/AmarTauqeer/Contract

[10] https://www.dalicc.net/

[11] https://wiki.ni4os.eu/index.php/License_Clearance_Tool_-_Description_and_Documentation

References

[1] C.L. Borgman, The conundrum of sharing research data, J. Am. Soc. Inf. Sci. Technol. 63 (6) (2012) 1059–1078, http://dx.doi.org/10.2139/ssrn.1869155.

[2] L. Lyon, Dealing with data: Roles, rights, responsibilities and relationships consultancy report, 2007.

[3] T. Cai, Y. Wu, H. Lin, Y. Cai, Blockchain-empowered big data sharing for internet of things, in: Research Anthology on Convergence of Blockchain, Internet of Things, and Security, IGI Global, 2023, pp. 278–290, http://dx.doi.org/10.4018/978-1-6684-7132-6.ch017.

[4] Q.H. Cao, G. Madhusudan, R. Farahbakhsh, N. Crespi, Usage control for data handling in smart cities, in: 2015 IEEE Global Communications Conference, GLOBECOM, IEEE, 2015, pp. 1–6, http://dx.doi.org/10.1109/GLOCOM.2014.7417270.

[5] N. Ahmed, R.A. Michelin, W. Xue, S. Ruj, R. Malaney, S.S. Kanhere, A. Seneviratne, W. Hu, H. Janicke, S.K. Jha, A survey of COVID-19 contact tracing apps, IEEE Access 8 (2020) 134577–134601, http://dx.doi.org/10.1109/ACCESS.2020.3010226.

[6] R.M. Savithramma, B.P. Ashwini, R. Sumathi, Smart mobility implementation in smart cities: A comprehensive review on state-of-art technologies, in: 2022 4th International Conference on Smart Systems and Inventive Technology, ICSSIT, 2022, pp. 10–17, http://dx.doi.org/10.1109/ICSSIT53264.2022.9716288.

[7] R.N. Zaeem, K.S. Barber, The effect of the GDPR on privacy policies: Recent progress and future promise, ACM Trans. Manage. Inf. Syst. 12 (1) (2020) http://dx.doi.org/10.1145/3389685.

[8] "GDPR", General data protection regulation (GDPR), 2018, Available online: https://gdpr.eu/what-is-gdpr/, (Accessed on 20 July 2022).

[9] A. Tauqeer, A. Kurteva, T.R. Chhetri, A. Ahmeti, A. Fensel, Automated GDPR contract compliance verification using knowledge graphs, Information 13 (10) (2022) http://dx.doi.org/10.3390/info13100447.

[10] T.R. Chhetri, A. Kurteva, R.J. DeLong, R. Hilscher, K. Korte, A. Fensel, Data protection by design tool for automated GDPR compliance verification based on semantically modeled informed consent, Sensors 22 (7) (2022) 2763.

[11] IAPP, IAPP-FTI, 2023, Available online: https://iapp.org/media/pdf/resource_center/IAPP_FTIConsulting_2020PrivacyGovernanceReport.pdf, (Accessed on 11 August 2023).

[12] D. Fensel, U. Şimşek, K. Angele, E. Huaman, E. Kärle, O. Panasiuk, I. Toma, J. Umbrich, A. Wahler, Introduction: What is a knowledge graph? in: Knowledge Graphs: Methodology, Tools and Selected Use Cases, Springer International Publishing, Cham, 2020, pp. 1–10, http://dx.doi.org/10.1007/978-3-030-37439-6_1.

[13] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G.d. Melo, C. Gutierrez, S. Kirrane, J.E.L. Gayo, R. Navigli, S. Neumaier, et al., Knowledge graphs, Synth. Lect. Data, Semant., Knowl. 12 (2) (2021) 1–257, http://dx.doi.org/10.1145/3447772.

[14] The smashHit project, UC1 - insurance services, 2020, Available online: https://smashhit.eu/d6-5-demonstrator-of-services-using-integrated-cpp-and-insurance-data/, (Accessed on 9 July 2022).

[15] The smashHit project, UC2 - smart city services, 2020, Available online: https://smashhit.eu/d7-5-demonstrator-of-services-using-integrated-traffic-smart-city-and-cpp-data/, (Accessed on 9 July 2022).

[16] smashHit consortium, smashHit concept (white paper), White paper is part of the smashHit project deliverable D2.2 smashHit Methodology, 2022, http://dx.doi.org/10.5281/zenodo.7870318.

[17] smashHit consortium, smashHit user & developer guidelines data provider & data processor, user guide part of the smashhit project deliverable D2.2 smashHit Methodology, 2022, http://dx.doi.org/10.5281/zenodo.7870766.

[18] Ç.K. Koç, F. Özdemir, Z. Ödemiş Özger, Rivest-Shamir-adleman algorithm, in: Partially Homomorphic Encryption, Springer International Publishing, Cham, 2021, pp. 37–41, http://dx.doi.org/10.1007/978-3-030-87629-6_3.

[19] D. Selent, Advanced encryption standard, Rivier Acad. J. 6 (2) (2010) 1–14, https://www2.rivier.edu/journal/roaj-fall-2010/j455-selent-aes.pdf.

[20] A. Kurteva, T.R. Chhetri, A. Tauqeer, R. Hilscher, A. Fensel, K. Nagorny, A. Correia, A. Zilverberg, S. Schestakov, T. Funke, E. Demidova, The smashHitCore ontology for GDPR-compliant sensor data sharing in smart cities, Sensors 23 (13) (2023) http://dx.doi.org/10.3390/s23136188.

[21] smashHit consortium, D7.5 demonstrator of services using integrated traffic, smart city and CPP, 2022, http://dx.doi.org/10.5281/zenodo.7868230.

[22] smashHit consortium, D6.5 - demonstrator of services using integrated CPP and insurance data, 2022, http://dx.doi.org/10.5281/zenodo.7867998.

[23] MyData Global and KIRAHub, H3C event: Built for people - human-centric solutions for the built environment, 2024, Available online: https://oldwww.mydata.org/h3c-event-built-for-people-6-april-2022/. (Accessed on 30 March 2024).