Demonstration Paper

Secure and Transparent Data Sharing with *TrustShare*: A GDPR-Compliant Platform

Sven Rasmusen*
Fujitsu Technology Solutions S.A.
Luxembourg, Luxembourg
sven.rasmusen@fujitsu.com

Konstantina Pityanou
Department of Electrical and Computer Engineering, Hellenic Mediterranean University
Heraklion, Crete, Greece
kpityanou@hmu.gr

Dimitra Papatsaroucha
Department of Electrical and Computer Engineering, Hellenic Mediterranean University
Heraklion, Crete, Greece
dpapatsa@hmu.gr

Sofiane Lagraa
Fujitsu Technology Solutions S.A.
Luxembourg, Luxembourg
sofiane.lagraa@fujitsu.com

Moussa Ouedraogo
Fujitsu Technology Solutions S.A.
Luxembourg, Luxembourg
moussa.ouedraogo@fujitsu.com

Evangelos K. Markakis
Department of Electrical and Computer Engineering, Hellenic Mediterranean University
Heraklion, Crete, Greece
emarkakis@hmu.gr

ABSTRACT

This paper presents *TrustShare*, a platform designed to facilitate secure and transparent data sharing within the European Data Space, addressing challenges in managing GDPR-compliant agreements for secondary data use. *TrustShare* offers a backend and an intuitive interface for defining, managing, and enforcing data access constraints. It utilizes a data registration and search engine, combined with a knowledge graph, to handle agreements. This system incorporates trusted timestamps and digital signatures, enhancing the security and traceability of data agreements between the provider and consumer. A demonstration scenario illustrates *TrustShare*'s functionality in a healthcare context, showcasing data registration, agreement creation, and secure data access between data providers and consumers. The recorded demo of the platform, available at https://youtu.be/JZ5Sd4SxU3k, showcases only the data provider workflow for *TrustShare*, focusing on data registration and agreement creation.

1 INTRODUCTION

A data space in Europe is a decentralized framework that enhances the data ecosystem by enabling secure and trustworthy data sharing among diverse actors [1] [2]. It provides a structured environment where data providers and consumers can share and access data across various sectors, promoting the secondary use of data for innovative research and analysis. By integrating data ecosystems through common governance and regulatory standards such as the General Data Protection Regulation (GDPR), and technical mechanisms, data spaces such as GAIA-X [3] and EOSC [4] ensure high-quality, interoperable data is accessible across sectors like health and agriculture. This setup fosters interoperability and trust, driving advancements while maintaining compliance with EU laws and standards.

In the data ecosystem, secondary use connects data providers and consumers by repurposing data beyond its original intent. Data providers collect and supply data through methods such as surveys or sensors, while data consumers utilize these data for analysis or decision-making. This interaction enables the flow and application of information across different domains [7]. However, current data sharing platforms for the secondary use lack a comprehensive and efficient solution to manage and monitor GDPR-compliant data sharing and agreements between data providers and consumers, particularly in complex scenarios with multiple stakeholders and diverse data types. Resulting in difficulties with the registration, search and effective enforcement of the data usage restrictions [6].

This paper addresses the challenges of ethical and secure data sharing within a European data space [5], particularly concerning secondary use of anonymized data [3] [4]. The problem is the lack of comprehensive, GDPR-compliant solutions for managing the data sharing with their agreements that account for data provider constraints, dynamic contexts, and the specific needs of secondary data utilization. Existing methods are insufficient, hindering innovation, collaboration, and equitable access to research, especially for under-resourced researchers. For instance, in the current landscape, a small research institution that seeks health data from multiple European hospitals for studies of rare diseases faces significant challenges due to fragmented data sharing processes [2]. Each hospital's unique data access restrictions and legal requirements lead to a cumbersome negotiation process, delaying research, and risking GDPR non-compliance. This example underscores the need for a comprehensive solution to streamline and enforce GDPR-compliant data sharing agreements efficiently [1].

In this paper, we propose *TrustShare*, a solution for secure and transparent data sharing by facilitating the creation, management, and enforcement of GDPR-compliant agreements between data providers and consumers. It offers a user-friendly interface for data providers to define data access constraints and for data consumers to accept and adhere to those terms, leveraging a knowledge graph for agreement management and incorporating trusted timestamps and digital signatures to ensure security and auditability. *TrustShare* has the following components:

*Both authors contributed equally to this research.
[1] https://wikis.ec.europa.eu/display/jrcdataspaceswiki/1.3++Definitions
[2] https://language-data-space.ec.europa.eu/help/faq/what-data-space_en
[3] https://gaia-x.eu/
[4] https://eosc-portal.eu/

- A user-friendly data sharing and agreement platform to manage GDPR-compliant data sharing agreements throughout their lifecycle.
- A knowledge graph to build agreement models from legal domains while ensuring GDPR compliance.
- Security and transparency through unique agreement hashes, trusted timestamps, and digital signatures.

A recorded demo highlighting a specific functionality: the data provider workflow, including data registration and agreement management, is available at https://youtu.be/JZ5Sd4SxU3k.

2 *TRUSTSHARE*: ENHANCING TRUST, SECURE, AND COMPLIANT DATA SHARING

In this section, we introduce the *TrustShare* platform, designed to enhance trust whilst ensuring secure and compliant data sharing.

2.1 Overview

Figure 1 shows the architecture of *TrustShare*. *TrustShare* consists of a data description, registration, and search module, along with an agreement management module.

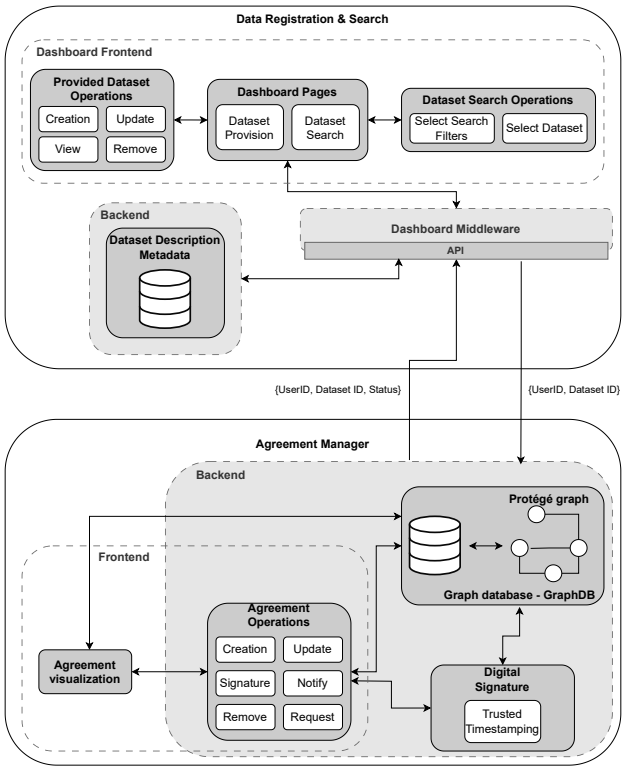

Figure 1: *TrustShare* architecture.

2.2 Data description & registration

The *TrustShare* Dashboard enables dataset registration with a front-end module for visualization. Data providers use a description form to submit metadata without disclosing actual data, ensuring compliance until agreements are signed. Providers can create, view, update, and delete dataset descriptions. The dynamic registration form records various details of the dataset, including general information, access, metadata schema, and structure. To enhance user experience, *TrustShare* maintains metadata and index catalog for dynamic form creation, allowing providers to select or add fields. New fields are saved for future use. After submitting the form, users go to the Agreement Manager to finalize the agreements.

2.3 Data search & request

The *TrustShare* Dashboard provides data consumers with the functionalities to search and explore descriptions of the registered datasets. Consumers can refine their searches using filters, improving efficiency. Privacy and security are maintained by requiring agreements prior to accessing datasets. Consumers create unique search queries using a dynamically updated attribute catalog, ensuring search criteria match current dataset descriptions. After submitting a query, consumers receive a list of matching datasets with descriptions. They can then select datasets and proceed to sign their relevant agreements to gain access.

2.4 Agreement Manager

The Agreement Manager facilitates the complete lifecycle management of GDPR-compliant data-sharing agreements. It handles the creation and removal of agreements, incorporating digital signatures and trusted timestamps to ensure authenticity and integrity. The manager maintains multiple agreement versions for auditing purposes and integrates with other modules (e.g., data manager, context handler) to coordinate agreement and data lifecycles. Its functionality is underpinned by the *Ontogreement* ontology [5], enabling semantic representation and automated checks for GDPR compliance. The following sub-sections outline and describe the components of the agreement manager.

2.4.1 Agreement Data Model. The ontology, modeled using Protégé, is exported as an RDF schema and stored in the GraphDB [6] graph database for practical use and interaction with other components of the system. This allows for maintaining the ontology's structure and enabling updates. The backend utilizes SPARQL for database queries. A data model encompassing eight GDPR-compliant objects (including Agreement, Purpose, Action, Signature, and User) facilitates data management and rule-based validation of agreement validity and data merging possibilities.

2.4.2 Agreement Operations. The *TrustShare* provides data providers with the ability to create and remove agreements, while data consumers can only sign the requested agreements. However, agreement removal is restricted to unsigned agreements.

2.4.3 Agreement digital signature based on Trusted Timestamping. To ensure the reliability and verifiability of digital signatures, the *TrustShare* employs a trusted timestamping module using Digital Signature Services (DSS[7]) compliant with eIDAS regulations. This generates cryptographic proof embedded within the agreement, preventing timestamp manipulation and accurately tracking its lifecycle. When a data provider creates an agreement, it will be signed by multiple data consumers who request the data to be used. In such cases, the data provider creates an agreement and signs it. The rest of the signatures given by the data consumers are in parallel after the data provider's signature as illustrated in Figure 2. Therefore, the agreement involves only

[5] https://github.com/FujitsuLUXECS/Ontogreement
[6] https://graphdb.ontotext.com/
[7] https://ec.europa.eu/digital-building-blocks/wikis/display/DIGITAL/Digital+Signature+Service+-++DSS

bilateral signatures between the data provider and the consumer, establishing a 1-to-1 relationship.

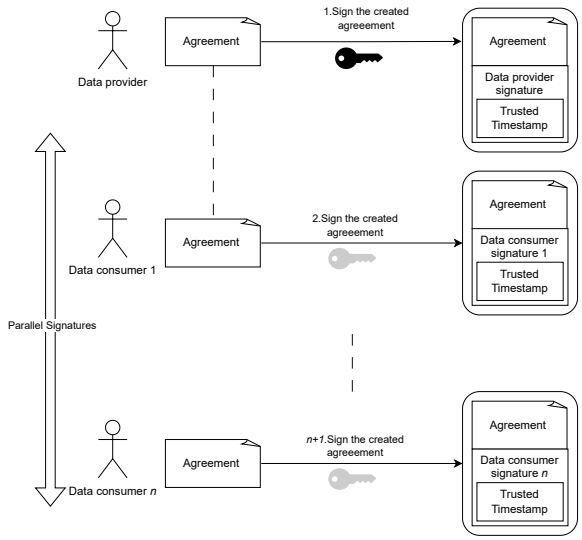

Figure 2: Parallel data agreement signatures based on Digital Signature Services (DSS)

2.4.4 Agreement visualization. Effective user interface design for both the data provider and the consumer is crucial to present complex data-sharing agreements in a clear and understandable manner. To address user comprehension challenges regarding privacy implications, *TrustShare* includes a user interface designed to promote careful consideration of the agreement terms before consent is given. The interface prioritizes clarity and avoids overly simplified representations.

2.4.5 Dataset and agreement management. *TrustShare* automatically generates Universally Unique Identifiers (UUIDs) for agreements. When a data provider wants to register a dataset, each dataset is directly linked to a dataset sharing agreement created during the dataset registration.

3 DEMONSTRATION SCENARIO

The demonstration scenario illustrates *TrustShare*'s practical application in secure, GDPR-compliant data sharing. Using a healthcare use case, it shows how data providers and consumers manage data agreements efficiently, ensuring compliance, and promoting secondary data use for research. The key functionalities highlighted include data registration, agreement creation, data search, and data agreement signature.

3.1 Use case: Healthcare

In this healthcare use case, a hospital (data provider) shares anonymized Covid-19 patient data with research institutions (data consumers) to study a new treatment. The demo presents four scenarios of data sharing and agreement management, facilitating secondary use of health data for both providers and consumers.

3.2 Data Provider Scenarios

In this demo, the hospital uses the data registration and agreement manager to register the dataset through metadata. It establishes a data sharing agreement that details the data, its use, access permissions, and duration. The hospital digitally signs the agreement with a trusted timestamp to prevent alterations.

Scenario #1: Data description & registration. The dataset registration form, shown in Figure 3, helps to submit detailed dataset information, ensuring that each dataset is well described to data consumers. It includes the dataset's name, domain, size, number of records, and an access endpoint. Data providers must also provide a brief description and select an access duration, specifying start and end dates for availability. In addition, the form has sections for metadata fields and dataset indices, which list column names and data types (e.g. integer, string). After completing the details, the data providers can submit the form and proceed to create and sign the agreement through the Agreement Manager.

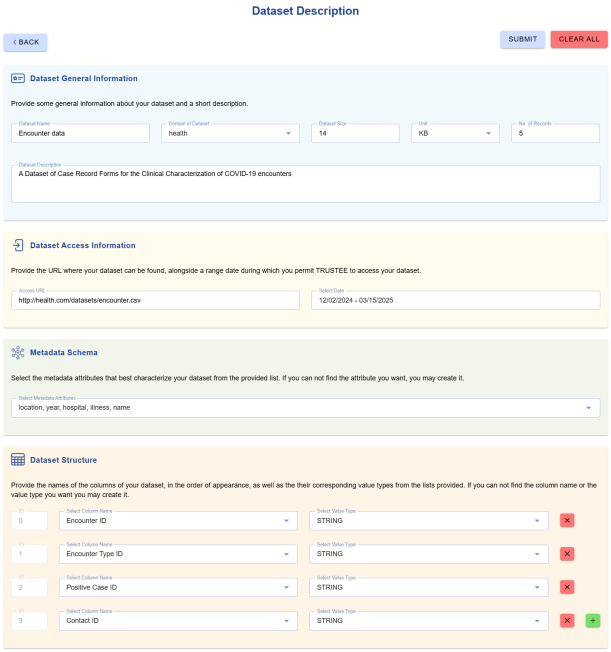

Figure 3: Dataset description & registration

Scenario #2: Agreement creation. The creation of agreement ensures that the data provider has control over how their data is used and shared. Following data registration, the data provider completes the agreement form by specifying various conditions, such as who can access the data, for what purposes, and any restrictions on data usage, as illustrated in Figure 4. The signature acts as a digital validation, confirming the provider's approval and intention to bind the agreement to the specific dataset. This step is crucial to maintain data governance and ensure compliance with legal and ethical standards. Once the agreement is finalized, it can be shared with potential data consumers, who must adhere to the specified conditions to access and use the data.

3.3 Data Consumer Scenarios

A research institution (data consumer) searches for and requests hospital data via the data search engine. It receives a data sharing agreement to access anonymized Covid-19 patient data for a treatment study. The institution reviews and digitally signs the agreement, then securely receives the data, adheres to the terms of agreement, and maintains confidentiality and security.

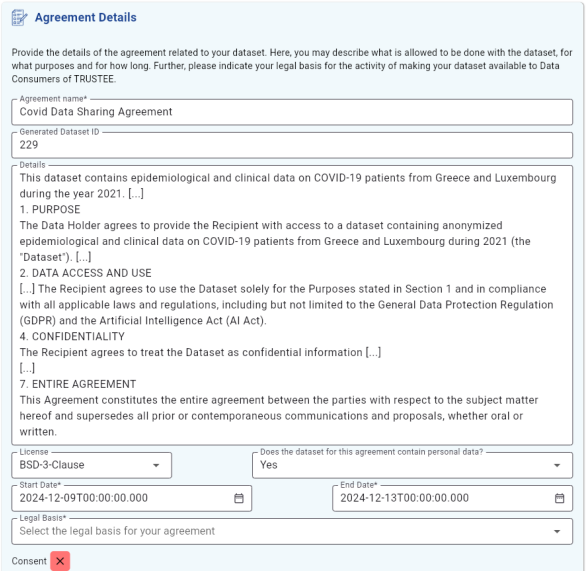

Figure 4: Data provider agreement creation and signing.

Scenario #3: Data search & request. Figure 5 shows that data consumers can formulate search queries on the *TrustShare* Dashboard using either the Build Query or Write Query features. The Build Query feature allows users to select filters and logical operations (e.g., "AND," "OR," "is equal to," "not equal to"). The Write Query feature lets users input queries directly into a code editor with formatting guidance. Consumers can filter by criteria such as the domain of the dataset, creation year, or specific properties (e.g., relevant illness). These filters are based on the attributes catalog of registered datasets, ensuring a comprehensive search experience. After constructing a query, users submit it to receive a list of matching datasets accompanied by their descriptions, then they select any desired datasets and proceed to sign agreements to unlock access to the data.

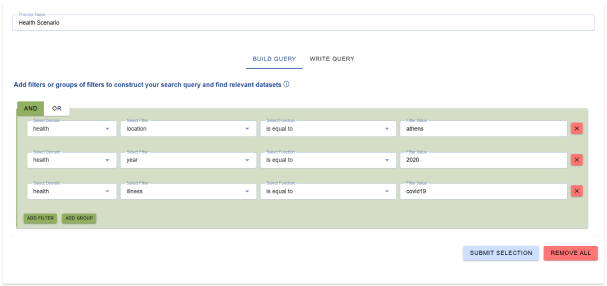

Figure 5: Dataset search & request (Build Query option)

Scenario #4: Agreement signature. After identifying and requesting a desired dataset, the data consumer thoroughly examines the attached agreement. The consumer carefully reviews the conditions and restrictions governing the data's use, ensuring a comprehensive understanding before providing its signature to confirm acceptance. The agreement's specifics are meticulously outlined and mutually signed by both parties, fostering clarity and mutual understanding, as illustrated in Figure 6. The digital signatures of both the hospital and the research institution are securely stored in a graph database. These signatures serve as verifiable evidence of the contract and facilitate auditing and compliance checks, reinforcing the integrity and accountability of the data usage agreement.

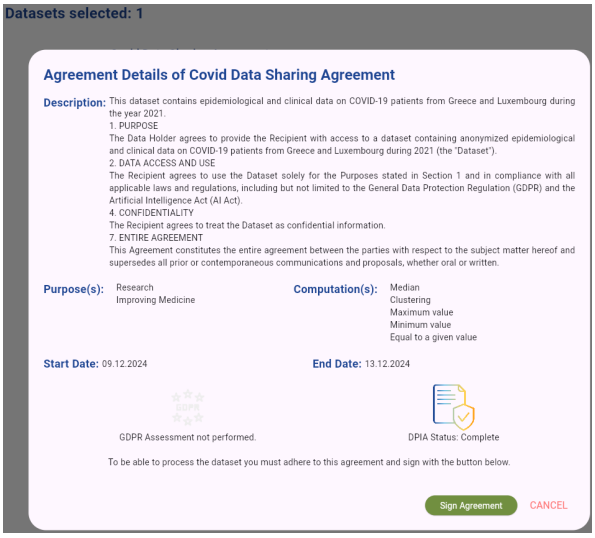

Figure 6: Data consumer agreement signature.

4 CONCLUSION

TrustShare effectively addresses the complexities of GDPR-compliant data sharing by providing a secure and transparent platform for managing datasets and their agreements, thereby facilitating the secondary use of data in research while ensuring compliance with legal and ethical standards. The future developments will focus on enhancing *TrustShare*'s scalability and integration capabilities to support a broader range of data types and sectors, further promoting interoperability and innovation in data-driven research.

ACKNOWLEDGMENTS

This work was supported by European Union's HE TRUSTEE project under the grant agreement number 101070214.

REFERENCES

[1] Pinky Bai, Sushil Kumar, Kirshna Kumar, Omprakash Kaiwartya, Mufti Mahmud, and Jaime Lloret. 2022. GDPR Compliant Data Storage and Sharing in Smart Healthcare System: A Blockchain-Based Solution. *Electronics* (2022). https://doi.org/10.3390/electronics11203311

[2] Sylvie Courbier, Rebecca Dimond, and Vicky Bros-Facer. 2019. Share and protect our health data: an evidence based approach to rare disease patients' perspectives on data sharing and data protection - quantitative survey and recommendations. *Orphanet Journal of Rare Diseases* (2019). https://doi.org/10.1186/s13023-019-1123-4

[3] Shona Kalkman, Menno Mostert, Christoph Gerlinger, Johannes JM van Delden, and Ghislaine JMW van Thiel. 2019. Responsible data sharing in international health research: a systematic review of principles and norms. *BMC Medical Ethics* (2019).

[4] Marlies Saelaert, Louise Mathieu, Wannes Van Hoof, and Brecht Devleesschauwer. 2023. Expanding citizen engagement in the secondary use of health data: an opportunity for national health data access bodies to realise the intentions of the European Health Data Space. *Archives of Public Health* (2023).

[5] Simon Scerri, Tuomo Tuikka, Irene Lopez de Vallejo, and Edward Curry. 2022. Common European Data Spaces: Challenges and Opportunities. In *Data Spaces*. Springer.

[6] Tsaone Tamuhla, Eddie T Lulamba, Themba Mutemaringa, and Nicki Tiffin. 2023. Multiple modes of data sharing can facilitate secondary use of sensitive health data for research. *BMJ Global Health* (2023). https://doi.org/10.1136/bmjgh-2023-013092

[7] Carol Tenopir, Natalie M. Rice, Suzie Allard, Lynn Baird, Josh Borycz, Lisa Christian, Bruce Grant, Robert Olendorf, and Robert J. Sandusky. 2020. Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLOS ONE* (2020). https://doi.org/10.1371/journal.pone.0229003