Towards Semantic Representation of Dataspace Technology Platform Architecture

Alexey (Aliaksei) Andrushevich
"Intelligent Semantic Systems" LLC
Minsk, Belarus
info@semantic.by
https://orcid.org/0000-0001-7138-8872

Iosif Vojteshenko, Matvey Kvetko
Belarusian State University
Minsk, Belarus
voit@bsu.by
https://orcid.org/0000-0002-0134-1793

***Abstract*—This paper explores the architectural principles and semantic aspects of building data space platforms with a focus on the MobiSpaces project centered on mobility and personal data protection. The paper describes the main technological components of the IDS RAM reference architecture, including data connectors, authentication, and certification services. Solutions for implementing an access security model, such as annotation of personally identifiable information and attribute-based access control (ABAC), are presented. Special attention is paid to state-of-the-art authentication techniques, including passwordless authentication using the FIDO2 standard and the use of electronic identity documents (eIDs), which contribute to achieving a high degree of trust and interoperability within data spaces. A key contribution of the research is the use of semantic technologies (in particular, the SCn/SCg OSTIS language) to describe architecture and access control, which increases the flexibility and context awareness of systems when exchanging data. The combined application of FIDO2 and ABAC is seen as a promising direction in the development of scalable and secure digital ecosystems.**

***Keywords*—Data spaces, Data Space Protocol (DSP), semantic representation, technical components of the dataspace architecture, FIDO2 specification, eID card, FIDO-AC framework**

I. Introduction

In an increasingly data-driven world, the ability to share information across organizational and sectoral boundaries is essential for innovation, efficiency, and social progress. However, traditional models of data sharing often compromise privacy, control, and interoperability. Data spaces have emerged as a novel approach to address these concerns by creating decentralized, interoperable environments for secure, and sovereign data exchange.

At their core, data spaces are ecosystems where multiple stakeholders -such as businesses, governments, and research institutions – can share data under commonly agreed upon rules and standards. Unlike centralized data platforms, data spaces emphasize data sovereignty, meaning that data owners retain full control over how their data aree accessed and used. This is achieved through mechanisms such as usage policies, access control, and trust frameworks, often supported by technologies such as decentralized identity management and secure data connectors.

One of the key drivers behind the development of data spaces is the growing recognition of data as a strategic asset. Data Spaces seek to empower sectors like healthcare, manufacturing, energy, and mobility by fostering cross-domain collaboration without sacrificing data governance.

Implementing data spaces, however, presents several challenges. Technical issues such as standardizing data models and ensuring interoperability between different IT systems must be addressed. Legal and ethical considerations, including compliance with data protection regulations like the General Data Protection Regulation (GDPR), are equally critical. Moreover, building trust among participants requires robust governance frameworks, clear contractual agreements, and transparent accountability mechanisms.

Despite these challenges, the potential benefits of data spaces are substantial. By enabling secure, scalable, and equitable data sharing, data spaces can drive innovation, reduce inefficiencies, and support the development of data-driven services and products. They also represent a step forward in achieving digital sovereignty, especially in contexts where data localization and regulatory compliance are paramount.

II. Semantic representation of concepts and components of data spaces

The SCn code (Semantic Code natural) and SCg code (Semantic Code graphical) of OSTIS technology [1] will be used to semantically represent dataspace concepts, standards, protocols and components. The materials published in [2] are used here.

Data Spaces
:= [An ecosystem of data models, datasets, ontologies, data exchange contracts and specialized management services provided by data centers, repositories, repositories, etc., and related soft

competencies (governance, social interactions, business processes)]

:= [Federation of decentralized data ecosystems through the use of interoperable software components for data connectivity without the need for central governance]

:= [A large-scale heterogeneous collection of data distributed across multiple sources in multiple formats, with a mechanism for handling structured, semi-structured, and unstructured data]

Data Space Protocol (DSP)

:= [Designed for the seamless exchange of data by clearly defining procedures and rules for publishing, negotiating, and accessing data across multiple platforms and systems. DSP is not yet an ISO standard]

Technical components of the IDS RAM dataspace reference architecture

:= [IDS RAM defines the key components, their interactions, and the principles that govern the architecture of data spaces]

⇒ *Partitioning the class of data space technical components by functional purpose**:

= {• *data space connectors*
  ⊃ *provides connectivity between data sources, manages data source metadata and data usage conditions, and sends or receives data based on its usage conditions*
• *authentication provider*
  ⊃ *is responsible for maintaining and validating the authentication of connectors (or more specifically, the participants themselves). The connector must be authenticated prior to performing any operations in the data space*
• *Clearinghouse*
  ⊃ *tracks all transactions; if the data is payable, the clearing center shares the information for settlement with the involved connectors.*

}

Authentication provider

⇒ *Partitioning **:

= {• *Certification Authority (CA – Certification Authority*
  ⊃ *responsible for issuing and revoking credential certificates*
• *Participant Information Service (ParIS – Participant Information Service)*
  ⊃ *accepts, stores, and provides descriptions of connectors, so that other connectors can query and retrieve information about participant credentials*
• *Dynamic Attribute Provisioning Service (DAPS)*
  ⊃ *allows additional attributes (such as temporary changes in participant trust, information about known vulnerabilities or new versions of software components in use, and certificate revocation) to be added to the connector description and verified*

}

The connectors are logically divided into two parts, shown in Figure 1: the control plane and the data plane. The control plane deals with data management (user identification, implementation of access and usage policies), routing and processing, while the data plane is responsible for data exchange.

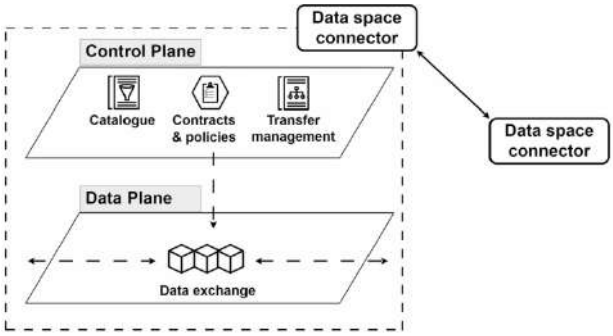

Figure 1. Connector: Control Plane and Data Plane [**?**].

In the context of the MobiSpaces project itself, the principles of data processing have been formulated [3], [4]:

- The platform must process only the minimum amount of personal data necessary to fulfill the specified purposes.

- The platform must have a legal basis for processing personal data.
- The platform must include mechanisms to enable data subjects to exercise their rights.
- The platform must ensure the accuracy and relevance of the personal data processed.
- The platform should exclude indefinite storage of personal data.
- The Platform shall implement measures to protect the confidentiality, integrity and availability of personal data, as well as ensure secure data processing using appropriate technical and organizational measures.
- The Platform shall keep records of personal data processing operations in order to demonstrate responsibility for the processing of personal data.

Components of the Data Spaces data access security model developed in the MobiSpeces project [3], [4]:

- Security Risk Modeler (SRM): This tool identifies security threats and risks to the system at the design level, based on the risk assessment method outlined in ISO 27005.
- Personally Identifiable Information (PII) Annotator: This component identifies data as PII or non-PII, preventing the transfer of personal data outside the system.
- Compliance Auditing Engine (CAE – Compliance Auditing Engine): This component is a distributed solution that processes the security pipeline in real time and includes monitoring agents.
- Protected Query Engine (PQE): This MobiSpaces component is implemented as a privacy-preserving algorithm that aims to protect data flows and provide query capabilities. It utilizes full homomorphic cryptography (FHE) for federated data sources.
- Attribute-Based Access Control (ABAC) mechanism: ABAC is a security model that uses attributes to grant access to MobiSpaces data and infrastructure. It can handle access control policy in a distributed and decentralized system.

III. Possible approaches to securing the use of data spaces

Data spaces consisting of multiple interacting information resources, software components and agents are characterized by high complexity in the context of user identity and privilege setting requirements. Organization of an effective and secure mechanism for authentication and authorization of users is a priority.

The authentication system should be compatible with various operating systems and platforms, as well as applications and devices. Such a system should be able to easily integrate with other systems without requiring significant changes to their structure and code. This can help reduce infrastructure and maintenance costs, since different authentication components do not need to be developed and maintained for each individual system. On the other hand, the ability to use a single authentication apparatus to access different systems can improve security because it can be designed to prevent different attacks.

The authentication system is an important element of the security system, but it works in conjunction with other components.

FIDO2 specifications. One of the trends in information security is the process of transition to passwordless authentication methods, in particular, more and more attention is paid to FIDO2 specifications [5]. Its application provides resistance to phishing, intermediary and replay attacks through the use of asymmetric cryptography and the generation of a cryptographically random buffer of bytes at each request.

The main components of the FIDO2 specification are the W3C WebAuthn web authentication standard and the Client-to-Authenticator Protocol (CTAP). Other key components of FIDO2:

- Platform Dependent Authenticator. This is a software module, implemented either as a standalone application or at the operating system level, used for authentication on a single device.
- Platform-independent authenticator. This is an external physical device, not tied to a specific platform (operating system), used for authentication on multiple devices. Platform-dependent and platform-independent authenticators act as a secure object integrated into or attached to the user's endpoint device (e. g., Windows Hello as a platform-dependent authenticator) (e. g., a USB hardware token as a platform-independent authenticator).
- Client. The client is played by the operating system and web browser. The client acts as the interface between the authenticator and the relyiing party.
- Relyiing party. Includes a web server and a FIDO server. The verifying party provides access to the protected resource, accepts requests from the Web server or application to enroll and authenticate users.

Semantic notation (using SCn and SCg languages of OSTIS technology) to the development of applications with passwordless authentication according to the FIDO2 specification is presented in [6]. One of the important processes in passwordless authentication is the user key registration process (see fig. 2).

Attribute-Based Access Control, ABAC. The increasing demand for the ABAC model is due to its advantages.

An important advantage is the absence of restrictions on the complexity of access rules. For example, when configuring access to some resource of the system, we can require from the user only his age or position, and also take into account a large set of rules, such as,

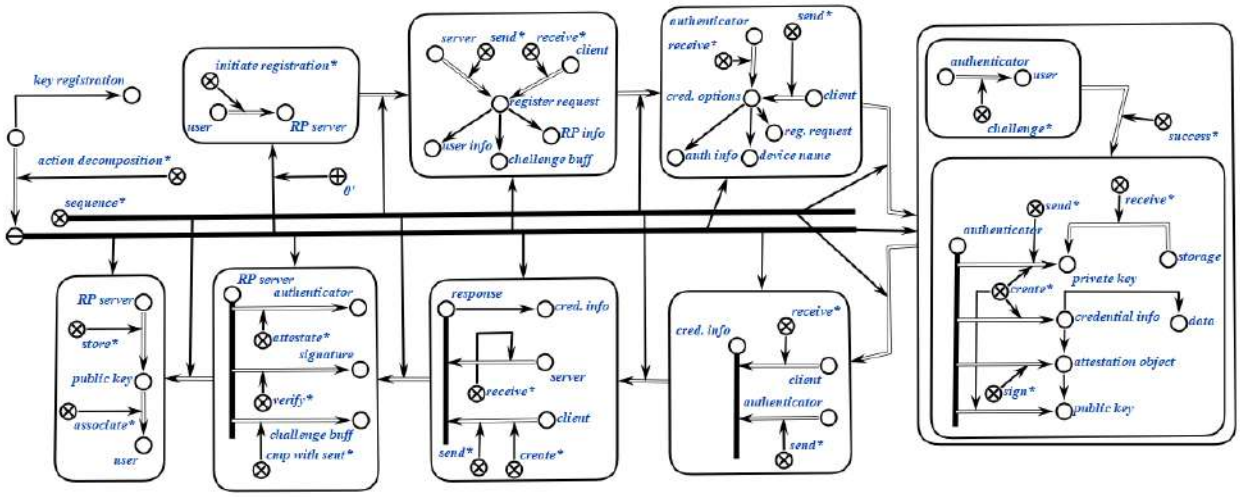

Figure 2. Key registration flow [6].

for example, department, specialization, project, level of experience or even citizenship. And for each resource the set of rules can be different. Support for dynamic attributes such as location or access time is provided.

The combined use of FIDO2 authentication and ABAC can significantly increase the level of flexibility and scalability of a security system. It is important to realize that these are two different processes, and the underlying FIDO2 specification (specifically WebAuthn) does not provide for such integration. Consequently, configuring ABAC and FIDO2 together may require additional effort and resources specific to each system or task.

The key issues here is which custom attribute medium to use. One proposal is to use the ABAC access model in conjunction with the OAuth 2.0 authorization protocol. This is a logical solution, but it does not provide a way to verify the user data provided by the service provider.

eID card. One option is the use of electronic identification documents (eID) storing the signature of the issuing party, the issuer.

The use of eID promotes interoperability on several levels: making it an effective and convenient means of confirming identity and sharing information in a digital environment.

One form of eID is a smart card with an embedded microchip. The chip contains an electronic means of biometric identification with personal data of the biometric document holder in accordance with the requirements of the International Civil Aviation Organization (ICAO) (fig. 3).

According to ICAO standards, to access the data presented on the electronic identifier the user needs to pass basic access control. For this purpose it is necessary to specify the number of the electronic document, expiration date, date of birth of the owner. After confirming the correctness of the data, the authorization system can receive user attributes. In order for the system to receive user attributes and perform all necessary checks, the application needs to be able to interact with the NFC-chip and extract data from the electronic document. All modern smartphones are equipped with an NFC-reader.

Basic access control is designed to ensure that card data can only be accessed when the card is physically accessed. Passive authentication proves that the contents of the Document Security Object and LDS are genuine and have not been altered. Active authentication provides secure information exchange and one-way chip authentication.

The ID-card of the Republic of Belarus can be used as an electronic biometric document complying with the standards set by ICAO.

Joint use of FIDO2 authentication, attribute authorization and eID card. In [8], the FIDO-AC framework is proposed, which is an extension of the basic FIDO2 specification. The essence of the obtained solution is to create an additional party - a mediator (mediator), which is responsible for validation of user data obtained, for example, from an eID-document (fig. 4).

A component released by Rock Solid Knowledge (RSK) for the ASP.NET Core 8.0 platform may be chosen as the library that allows the web application under development to act as a relying party. The FIDO2 library developed by RSK allows the web application to act as a WebAuthn validator by providing an IFidoAuthentication service containing methods for generating requests to a client that interacts with the WebAuthn API [9].

A semantic representation of the joint use of FIDO2 authentication, attribute authorization and eID card is given in [10].

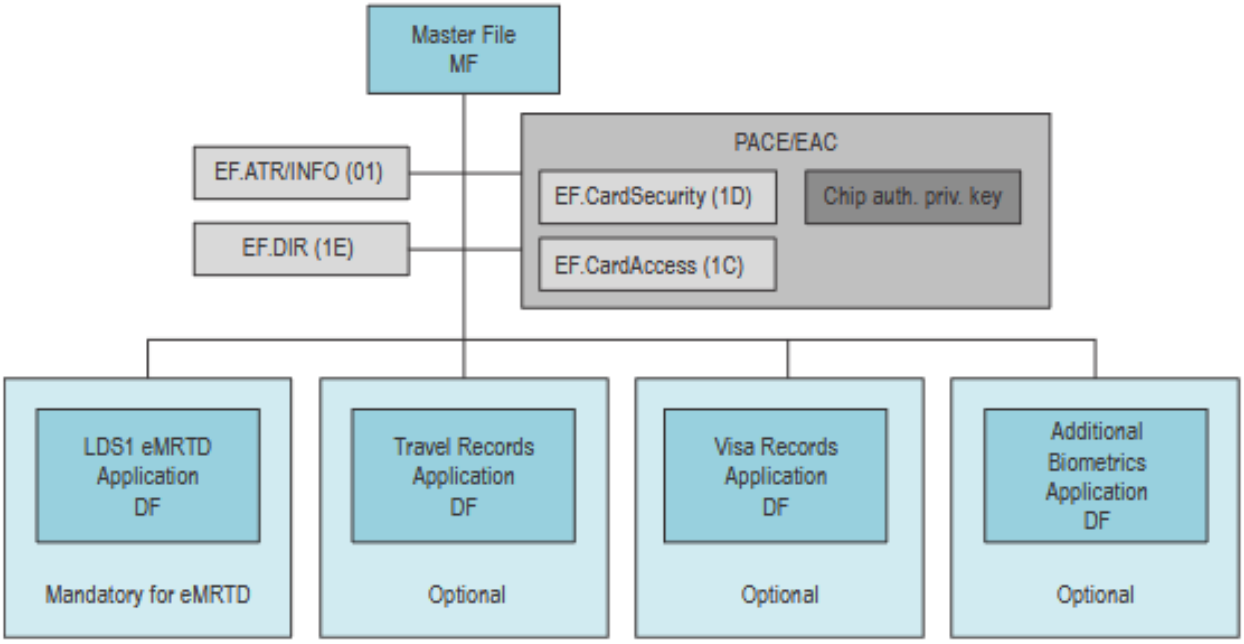

Figure 3. ICAO Logical Data Structure (LDS) [7].

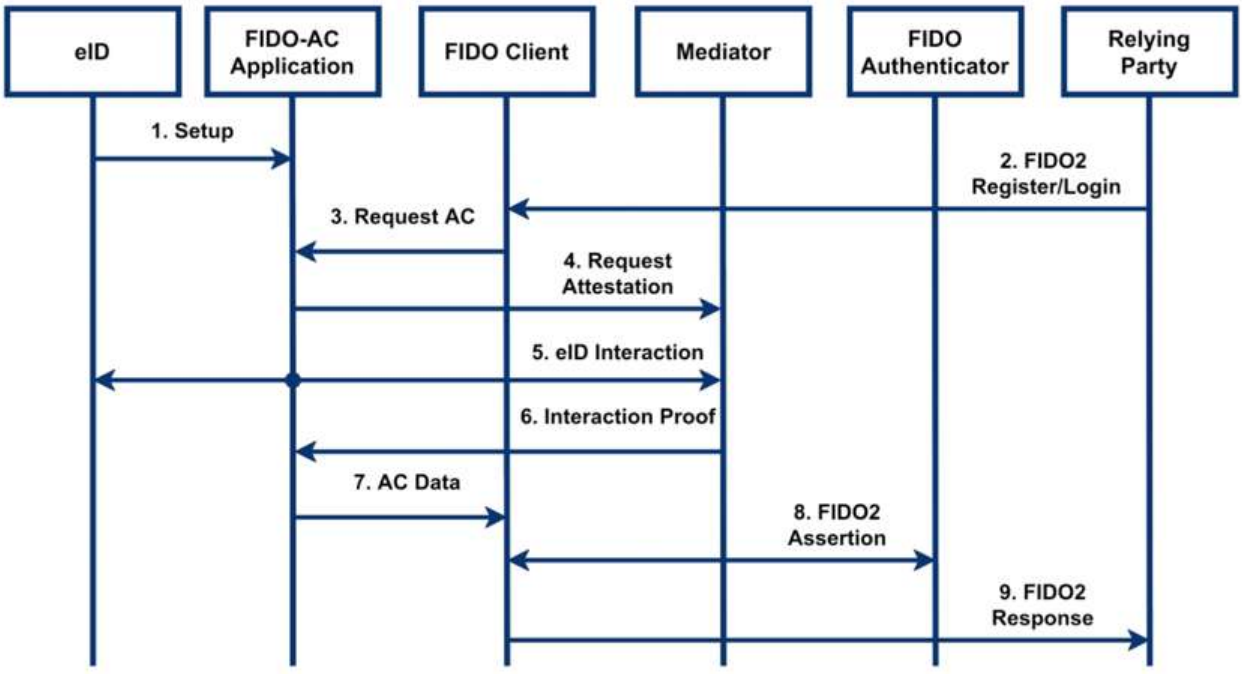

Figure 4. Interaction of FIDO-AC framework components.

IV. Conclusion

The advancement of data spaces as a paradigm for decentralized, secure, and sovereign data sharing presents a transformative opportunity for how data is accessed, managed, and governed across sectors. This research has explored the architectural underpinnings, semantic representations, and technical challenges involved in developing such platforms, with particular attention to initiatives like the MobiSpaces project. The integration of interoperable components such as data connectors, identity providers, and compliance engines illustrates the modular and federated nature of data space infrastructures, enabling a flexible yet robust framework for data collaboration.

A key contribution of this work lies in emphasizing the role of semantic technologies in enhancing interoperability and context-aware access control. By leveraging semantic representations—alongside advanced authentication (e. g., FIDO2) and authorization (e. g., ABAC) mechanisms—data spaces can support nuanced, policy-driven access to sensitive information while preserving privacy and user control. The combined use of electronic identification (eID) and secure cryptographic protocols further reinforces trust and transparency within these ecosystems.

Acknowledgment

The authors thank the teams of the Department of Software Engineering of the Belarusian State University and the Department of Intellectual Information Technologies of the Belarusian State University of Informatics and Radioelectronics for their help and valuable comments.

References

[1] V. V. Golenkov (ed.). Tehnologija kompleksnoj podderzhki zhiznennogo cikla semanticheski sovmestimyh intellektual'nyh komp'juternyh sistem novogo pokolenija. [Technology for Integrated Life Cycle Support of Semantically Compatible Intelligent Computer Systems of the New Generation], Minsk, Bestprint, 2023, 1064 p.

[2] M. Bacco, A. Kocian, St. Chessa, A. Crivello, P. Barsocchi. What are data spaces? Systematic survey and future outlook. Manlio Bacco at al. Available at: https://www.sciencedirect.com/science/article/pii/S2352340924009314.

[3] M. C. Compagnucci, Ih. Nwankwo, N. Masood, N. Mehandjiev, M. Fenwick. Data Protection and Data Security by Design in Mobility Data Spaces. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstractid=4936779.

[4] Ch. Doulkeridis et al. MobiSpaces: An Architecture for Energy-Efficient Data Spaces for Mobility Data, 2023 IEEE International Conference on Big Data (BigData), Sorrento, Italy, 2023, pp. 1487-1494.

[5] WebAuthn API specification. Available at: https://w3c.github.io/webauthn/

[6] A. Zhidovich, A. Lubenko, I. Vojteshenko and A. Andrushevich Semantic Approach to Designing Applications with Passwordless Authentication According to the FIDO2 Specification. Open Semantic Technologies for Intelligent Systems, 2023, Iss. 7, pp. 311–316.

[7] Doc 9303. ICAO. Available at: https://www.icao.int/publications/pages/publication.aspx?docnum=9303.

[8] Wei-Zhu Yeoh [et al.] Fast IDentity Online with Anonymous Credentials (FIDO-AC). Proceedings of the 32nd USENIX Security Symposium, 2023, P. 3029–3046

[9] RSK FIDO2 for ASP.NET documentation. Available at: https://www.identityserver.com/documentation/fido2.

[10] A. Zhidovich, A. Lubenko, I. Vojteshenko Semantic Notation of Access Control Technology based on eID Identification, FIDO2-Authentication and Attribute-Based Authorization in Digital Environment. Open Semantic Technologies for Intelligent Systems, 2024, Iss. 8, pp. 371–376.

К СЕМАНТИЧЕСКОМУ ПРЕДСТАВЛЕНИЮ АРХИТЕКТУРЫ ТЕХНОЛОГИЧЕСКОЙ ПЛАТФОРМЫ DATASPACE

Андрушевич А. А., Войтешенко И. С., Кветко М. В.

В данном исследовании изучаются архитектурные основы, семантические представления и технические проблемы, связанные с разработкой технологической платформы dataspace, при этом особое внимание уделяется таким инициативам, как проект MobiSpaces. Интеграция таких совместимых компонентов, как соединители данных, провайдеры идентификации и механизмы обеспечения соответствия, иллюстрирует модульную и федеративную природу инфраструктур пространства данных, позволяя создать гибкую, но надежную основу для совместной работы с данными.

Ключевой вклад данной работы заключается в подчеркивании роли семантических технологий в повышении функциональной совместимости и управлении доступом с учетом контекста. Благодаря использованию семантических представлений, а также усовершенствованных механизмов аутентификации (например, FIDO2) и авторизации (например, ABAC) пространства данных могут поддерживать тонкий, определяемый политикой доступ к важной информации, сохраняя при этом конфиденциальность и контроль пользователей. Совместное использование электронной идентификации (eID) и безопасных криптографических протоколов еще больше укрепляет доверие и прозрачность в этих экосистемах.

Received 31.03.2025