Igor KOROPIECKI, Krzysztof PIOTROWSKI
IHP – Leibniz Institute for High Performance Microelectronics
Im Technologiepark 25,
15236 Frankfurt (Oder), Germany

IOT AND DATA SPACES AS A HIGHLY DISTRIBUTED MEASUREMENT AND CONTROL SYSTEM

Ongoing digital transformation creates new challenges for interoperability between distributed measurement platforms. The European Union introduces data spaces, which are on their way to become the standard for trusted data exchange between stakeholders. This proliferation creates new opportunities for value generation through data, service and application markets. The paper discusses the challenges of adapting existing concepts to create solutions for the emerging ecosystem of data-driven transactions.

IoT i Data Spaces jako wysoce rozproszony system pomiarów i sterowania

Trwająca transformacja cyfrowa stwarza nowe wyzwania dla interoperacyjności między rozproszonymi platformami pomiarowymi. Unia Europejska jest w trakcie wprowadzania data spaces, czyli standard do zaufanej wymiany danych między interesariuszami. Tworzone są nowe możliwości generowania wartości poprzez rynki danych, usług i aplikacji. W artykule omawiane są wyzwania adaptacji istniejących systemów w celu stworzenia rozwiązań dla tworzonego ekosystemu transakcji opartych na danych.

1. INTRODUCTION

The technological landscape of today's world is filled with various systems that generate immense amounts of data. The rapid growth of IoT systems without proper standardization leads to creation of many systems with walled-off datasets.

Typical IoT system consists of several physical devices equipped with sensors, software, and technologies that exchange data with other devices and systems over one or more communication networks. The systems are built to automate tasks, optimize processes, and increase efficiency. Each IoT system collects valuable insights that can be used to make better decisions.

The repeating practice is custom implementations that involve approaches specifically tailored to the unique challenges of the problem being addressed. It leads to variations in data models, communication protocols, and access control mechanisms. Often, the interactions with other systems are also implemented differently and require specific knowledge. Most of the current approaches result in isolated data silos. It is understandable that stakeholders restrict access. However, with the continuously growing amounts of data, we are missing out on the benefits of interoperability and the ecosystem that it creates. Proper mechanisms can be created where (among others) data producers are compensated and data consumers have the necessary access.

The opportunity has been recognized by the European Union, which invested significant amounts of resources to create *data spaces* [1, 2], formed as a federated digital infrastructure for data-related exchange. It introduces a trusted data ecosystem that enables collaboration between stakeholders (public, private or personal), provides standardized mechanisms, models, and governance to bring order and enforce clear data control. It essentially creates a single market for data. Each sector (industry, health, finance, etc.) is a separate data space. New systems use the data space principles, language and rules in context of the data they incorporate, while existing systems are integrated through adapters. Such approach is convenient for fostering innovation and maintaining access to the existing infrastructure.

The landscape of the data spaces reference architectures is evolving. Specific reference architectures can vary based on sector needs. Typical implementation (of the architecture) provides a set of building blocks that can be used to create a system through using highly standardized vocabularies, models and data connectors. The creators still have the freedom in their deployment,

however they have to adhere to the architecture to be able to participate in one or more data spaces. Thanks to this concept, it is possible to create highly distributed measurement systems that understand outputs produced by each other. The Fig. 1 presents an example architecture, where many systems allow different producers and consumers to participate in the data space (and use its services) through data connectors, forming a single market within a data space, with possibility to extend to more spaces.

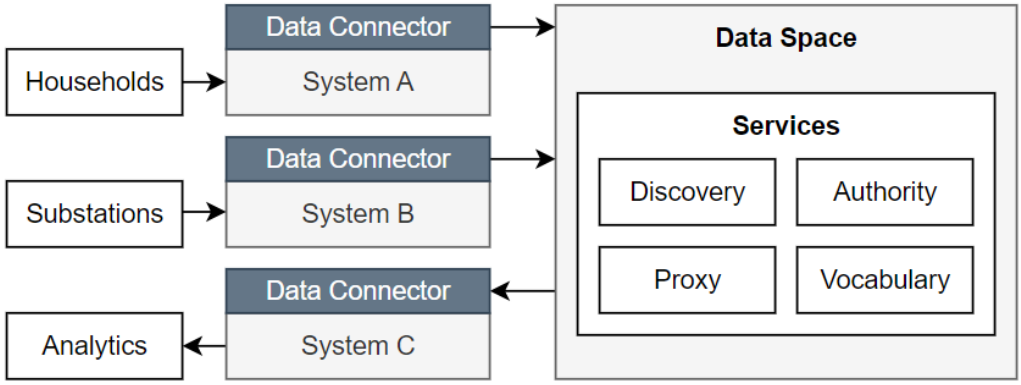

Fig. 1. Common example of data space architectures
Rys. Fig. 1. Powszechny przykład architectury data space

The adoption of the data spaces concept in all sectors of IoT could enable a greater synergy between systems. This would lead to large amounts of diverse data sources available for further implementations, which rely on large datasets (machine learning, decision making, etc.). Before such systems are created, several challenges must be addressed.

2. RELATED WORK

The European Union has set the general direction through legislative documents (European Data Strategy, Data & Governance Acts). The documents outline a vision for a European data space, fostering innovation, economic growth, and fair competition, while also ensuring data privacy and security.

The most notable initiative that contributes to this vision is the International Data Spaces Association (IDSA), which created the Reference Architecture Model (RAM) [3]. The model serves as the technical blueprint for building the secure and interoperable data spaces. It defines the roles of various participants (data providers, consumers, etc.), communication protocols for secure data exchange, and security mechanisms like access control and encryption.

Other notable initiatives focus on building secure and robust data handling frameworks and standards (e.g. Gaia-X, OpenDEI), while others provide technical solutions and infrastructure components (e.g. FIWARE, IDS Testbed). Additional initiatives focus on collaboration and business development within the ecosystem (e.g. DSBA, DSCC).

Also, there are many projects across several sectors that build on the foundation created by the previously mentioned (more general) initiatives. The European Commission is also working on Simpl [4], an open-source smart middleware platform that aims to enable cloud-to-edge federations through providing a highly modular and decentralized abstraction between the data and infrastructure layer. Simpl exposes platform capabilities as both centralized and decentralized components (services). Examples of centralized services include data catalogues, vocabulary providers or identity authorities. Decentralized components include Simpl Agents that reside on the premises of data, infrastructure and application providers and enable them to communicate on a P2P basis.

3. PROPOSED APPROACH

An example scenario of smart grid operations (metering, actuation, analytics, adaptation) is considered. In this scenario, interconnected devices (smart meters, computers, servers, etc.) are installed on user premises (households, companies), at critical infrastructure points (transformer stations, substations), and electricity company premises. The result is a data collection network, that captures insights on every stage of the energy supply chain, and allows to run various algorithms, which optimize operations and mitigate disruptions. The difference from current approach is the location of the data, which is not stored centrally. Instead, each entity has the capacity to store their own data, which enables high levels of privacy and control, and also aligns with the emerging adoption of edge paradigms in modern distributed systems. Stakeholders can use services provided by the data collection network to find data sources of interest and use payment mechanisms to compensate the source. Other benefits include processing closer to the data source (edge paradigm), which allows for a more distributed processing footprint and potentially faster access times.

The example scenario is a step towards creating privacy-by-design systems, addressing the growing concern of data safety and possible reluctance towards participation in novel smart grid operations, which inevitably collect large amounts of data that are considered private (or delicate). It is worth noting, that surveys [5, 6] have shown that users are gaining knowledge about such issues and are becoming increasingly aware of their data privacy rights. In the scenario, there are several challenges that must be addressed to successfully deploy a system compatible with data spaces.

Standardization, which is the crucial step for the success of new generation of IoT systems (and data spaces). Main goal is to eliminate fragmentation (e.g. incompatible models and data formats, lack of common communication protocols) resulting from the absence of common standards. Solutions include development of new standards and updates to existing ones, covering many areas (e.g. ontologies, data spaces, IoT, cybersecurity, privacy). This includes participation in standardization bodies (e.g. ISO, IEC, IDSA), aimed at defining reference architectures, interoperability profiles (e.g. standardized data formats, APIs), and trustworthiness.

Semantic and technological interoperability, which are required to ensure seamless communication and data exchange between systems and devices that have different implementations. It can be addressed through integration with existing platforms and connectors (e.g. IDSA), leveraging open catalogues (e.g. federated open service catalogue [7]) and building systems according to reference architectures that support data spaces.

Trust and sovereignty, which are required to establish a trusted space for data transactions in decentralized environments. The solution includes firm data usage control policies (e.g. access restrictions, encryption standards, fine-grained permissions) based on shareable data assets (e.g. customer profiles, sensor data) and defined stakeholder roles (e.g., administrators, analysts, end-users).

Maximizing data value and ensuring governance, which can be addressed through development of governance frameworks and policies (e.g. data access controls, data retention policies) to manage data usage and facilitate value creation, additionally combined with technology solutions (e.g. data analytics platforms, AI-driven tools) that further accelerate data value generation.

Security and privacy addresses storing sensitive user data on individual devices. Proper solutions include encryption, authentication and authorization, and compliance with regulations. Also, ensuring user privacy and obtaining consent for data usage becomes complex without a central governance framework.

Scalability and reliability poses a significant challenge in a distributed storage system across potentially millions of devices. Offline functionality adds more complexity, as data synchronization between devices might be disrupted. Solutions include distributed ledger technologies, microservices architecture, edge computing and redundancy and disaster recovery mechanisms.

The emergence of data markets introduces complexities regarding data ownership (e.g. user-generated content, health records), access rights, and pricing models (e.g. pay-per-use, subscriptions). Addressing these challenges requires agreeing on appropriate governance frameworks and regulatory mechanisms to allow fair competition.

4. CONCLUSIONS AND FURTHER STEPS

The European data spaces market is still forming. Some initiatives work on reference architectures, while some implement them to offer robust infrastructure for secure data exchange and thriving data marketplaces, however concrete market mechanisms are not yet established. The emergence of data markets introduces complexities regarding interoperability, data ownership, access rights, and pricing models. Addressing these challenges requires developing robust governance frameworks and regulatory mechanisms to foster fair competition and ensure data privacy and security. Common standards, technical infrastructure, and effective governance frameworks have the ability to transform the European data spaces into a connected ecosystem for open and secure data exchange.

Authors plan to address the mentioned challenges, while adopting a data exchange platform to align with the concept of data spaces.

REFERENCES

1. European Commission: Common European Data Spaces. Available at: https://digital-strategy.ec.europa.eu/en/policies/data-spaces (last accessed on 23.04.2024)
2. European Commission: Towards a common European data space. COM(2018) 232 final. April 2018.
3. IDSA: Reference Architecture Model 4. Available at: https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4 (last accessed 23.04.2024)
4. European Commission: Simpl. Available at: https://digital-strategy.ec.europa.eu/pl/policies/simpl (last accessed on 23.04.2024)
5. Cisco: Cisco 2023 Consumer Privacy Survey (Generation Privacy: Young Consumers Leading the Way). Available at: https://www.cisco.com/c/en/us/about/trust-center/consumer-privacy-survey.html?CCID=cc000742 (last accessed on 23.04.2024)
6. KPMG: Corporate data responsibility: Bridging the consumer trust gap. Available at: https://kpmg.com/us/en/articles/2023/bridging-the-trust-chasm.html (last accessed on 23.04.2024)
7. REFEDS: Service Catalogues in a Federated Context. Available at: https://refeds.org/wp-content/uploads/2018/10/ServiceCatalog-Evaluation.pdf (last accessed on 23.04.2024)