G0704

Interoperability by Sovereign and Secure Data Exchange in Trustworthy Data Spaces

Andreas Rumsch, Eugen Rodel, Christoph Imboden
Lucerne University of Applied Sciences and Arts, Horw/Switzerland;
Tel.: +41-41-349-3391
andreas.rumsch@hslu.ch

Abstract

More and more services need data from different sources. An example is a grid service aggregating distributed loads in a pool. Such a pool could consist of a large number of different loads, each from different manufacturers. The grid service then needs access to the data of all these devices. To keep costs low and to avoid the installation of gateways in the field, already existing infrastructure and data connections shall be used. This is where Data Spaces come into play. A Data Space builds a trustworthy platform to exchange data under full control of data owners. Open questions regarding the implementation of a Data Space are: Are software components available? How to ensure trustworthiness through governance? How to integrate existing standards? How can business cases benefit from Data Spaces?

Data Spaces set the concept and define the required elements, covering technical aspects as well as governance and legal topics. The authors have implemented a demonstrator of a Data Space, which shows the challenges one is facing when implementing or when connecting to a Data Space via so called connectors. Business cases have been developed in a series of workshops with representatives from industry.

The demonstrator is built based on a software framework provided by the International Data Spaces Association (IDSA). This framework shows good maturity compared to others. Based on the fundamentals of IDSA, it is possible to build a trustworthy Data Space, although the effort to integrate the connectors is still high and must be reduced. Many different ontologies representing standards already exist with very specific applications. The concept of Data Spaces eases the integration of ontologies via a vocabulary hub. Beside technical aspects, implementing Data Spaces also involves complying with data laws. The framework takes these into account and provides support.

Data Spaces can be seen as an open infrastructure for secure and sovereign data exchange, based on agreements, rules and standards. This is crucial for innovative business models. A common governance is crucial for gaining stakeholders' trust and willingness to participate in a Data Space that enable a future data economy. This necessitates a neutral and trustworthy non-profit organization to establish and operate the Data Space and the required infrastructure.

Further projects to implement Data Spaces have been initiated where grid services or a web-based HEMS will be realized to show the benefits of using Data Spaces.

Introduction

The expansion of PV systems in Europe is progressing. Since 2021 alone, installed capacity in the EU has risen from 164 GW to 260 GW in 2023 [1], and in Switzerland from 3.4 GW to 5.2 GW in the same period [2]. This is leading to an increasing decentralisation of electricity generation, as smaller and therefore more plants are producing this energy. Measures to stabilise the electricity grid must adapt to this change. One approach is to decentralise stabilisation by having many electrical loads such as charging stations, heat pumps, boilers, washing machines or dishwashers form a pool in which the individual appliances are switched on or off as required. Further, PV systems, batteries in houses and cars, or smart meters must be monitored. This requires a high degree of interoperability, which is an increasingly important requirement for stabilising the grid. In the example given, interoperability means that the operator of the pool can access data from different devices from different manufacturers and can also control these devices.

A classic approach to achieving interoperability is the use of gateways (fig. 1). The pool operator has access to the data of the devices via the gateways and can also control them in this way. The main task of the gateway is to standardise the protocols used by the various device manufacturers to provide a unified interface for the pool operators.

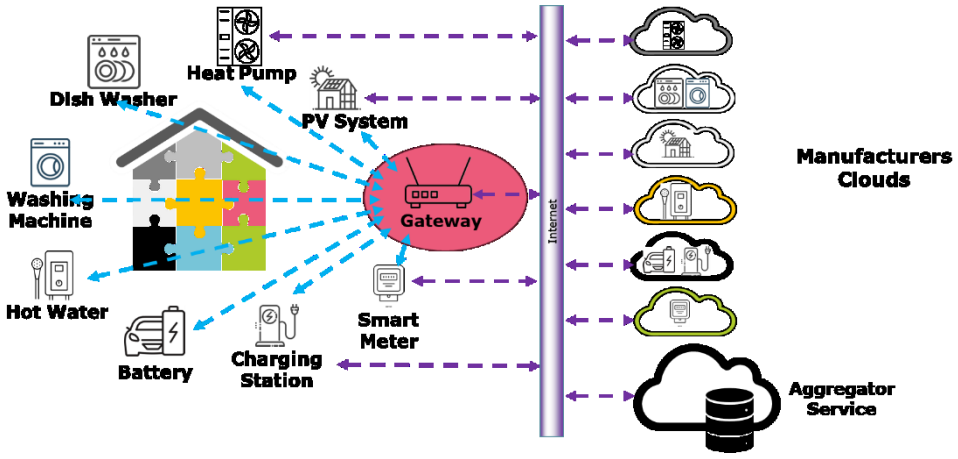

Fig. 1:
Classic approach to achieving interoperability with gateways.

A major disadvantage of the gateway approach is that a gateway must be installed in every household participating in the pool. In addition to the material costs for the gateway, the costs for installation and configuration are particularly significant. In our test area in Switzerland, we account for several hundred CHF per installation. These costs can make it impossible to roll out a good technical solution profitably on the market. Solutions that can do without the gateway would therefore have a clear advantage.

There are solutions without a gateway since modern devices are already equipped with an interface to the manufacturer's cloud. A pool operator could save costs by accessing the manufacturers' clouds directly (fig. 2). This is possible if the manufacturers grant access to the information stored in their cloud via an API.

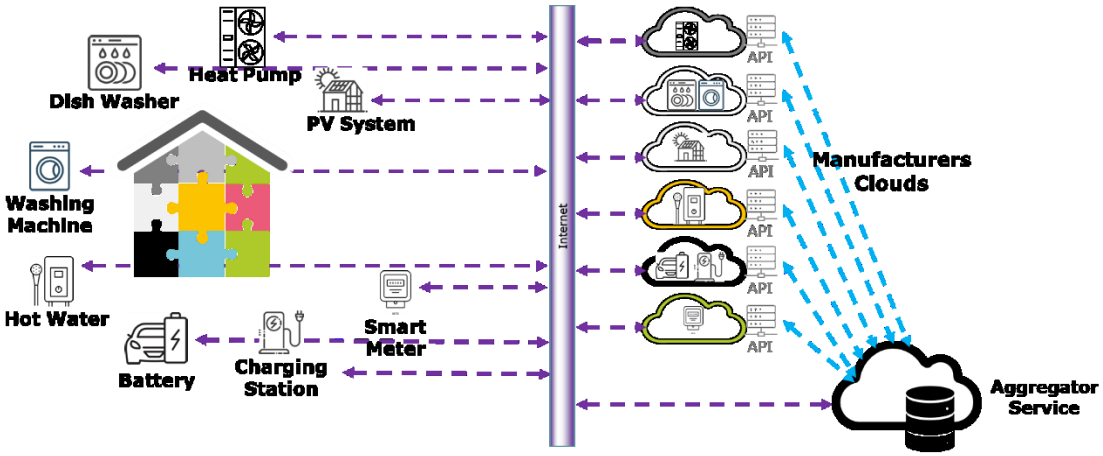

Fig. 2:
A pool operator can access the information stored in the manufacturers' clouds via APIs.

The approach of accessing the data via an API entails some disadvantages in terms of potential security risks. The manufacturers of the devices and the pool operator are responsible for ensuring that data security is maintained. And because a pool operator must agree with several manufacturers how he can access the data, the pool operator must implement appropriate security measures with all manufacturers, which in turn entails a high time and financial outlay. One of the questions for manufacturers is to what extent they can trust the pool operator to use the data in accordance with the agreement.

Another aspect is the extent to which the owners of the data (i.e. the users of the devices) have control over how their data is used. An agreement would have to be reached between the three parties of user, manufacturer and pool operator, which in turn can involve a great deal of effort.

This is where Data Spaces come into play. Only known and authorised organisations or individuals can participate in a Data Space. The Data Space thus represents a trustworthy infrastructure in which all participants can securely exchange data in accordance with the data owner's specifications. In addition to software components, a Data Space also includes a set of governance rules that cover processes at an organisational and legal level. Data exchange does not take place directly between providing and consuming participants, but rather the Data Space serves as an intermediary (fig. 3). The rules for data exchange are defined in the Data Space, which also eliminates the need for participants to reach a bilateral agreement.

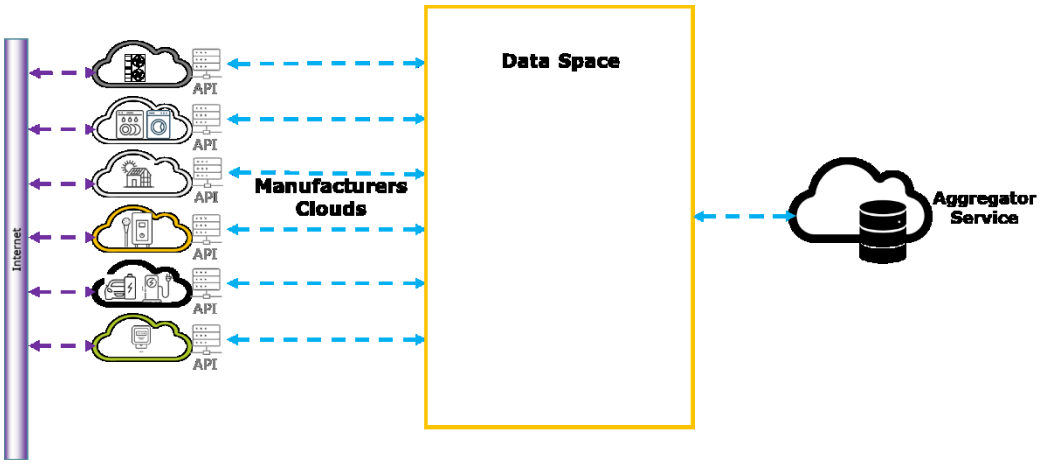

Fig. 3:
Data exchange takes place with the Data Space as an intermediary, controlled by a set of rules defined in the Data Space.

This paper describes what Data Spaces are, and how they are used in real world scenarios. Chapter 1 describes the concept of Data Spaces, chapter 2 shows in more detail how the example of pooling loads for flexibility is realized using the technology of Data Spaces. Chapter 3 dives into the realization of a Data Space and the challenges encountered. The benefits and the potential of using Data Spaces are discussed in chapter 4.

1. Concept of Data Spaces

This chapter delves into a general architecture overview, and a specific description of the software components utilized. Integral to Data Spaces are specific requirements:

– Data is not redundantly stored; it remains in its original location.
– Data owners retain control over determining access permissions for their data.
– Participants can trust the identity authenticity of others.
– No centralized entity manages data.

A Data Space is a data integration concept based on four pillars [3]:

– Data is left where it is created or managed, which results in a distributed data architecture and a distributed data infrastructure.
– There is no common database schema (data integration on the semantic level) required; therefore, vocabularies are needed to achieve semantic interoperability between the distributed data.
– Data Spaces allow for data redundancies and «co-existence» of data.
– Data Spaces can be nested, i.e., several Data Spaces can overlap but do not have to be disjoint and participants can be part of multiple Data Spaces.

In addition, the concept of Data Spaces comprises data sovereignty and data traceability. Data providers want to determine who can do what with their data, and they want transparency about what happens to their data when they share it. Further, data recipients must be able to trust that the data providers are who they say they are. A trust anchor is needed and is achieved through a certification process. A digital certificate is required to document this trust anchor.

The foundation of Data Spaces lies in a distributed software architecture. Connectors play a pivotal role in facilitating the provision and utilization of data while upholding data sovereignty. These connectors serve as the distributed data endpoints. Additional components are essential to enable these endpoints to discover each other and engage effectively. Broker services act as intermediaries, connecting data providers with data consumers and facilitating with the help of a vocabulary hub data exchange, ensuring a harmonious match between data supply and demand. Clearinghouse services oversee the successful execution of data transactions without delving into the data itself. To ensure trust, the identity provider only grants access to the data space to participants with a valid certificate. The components mentioned here are described in fig. 4.

In summary, Data Spaces unlock the potential for innovative business models by bridging the gap between data integration and application-driven value creation. Through a distributed software architecture and the careful orchestration of connectors, broker services, and clearinghouse services, Data Spaces empower organizations to harness data while respecting the sovereignty of its providers, thus catalysing a new era of data-driven innovation.

Fig. 4:
Components needed for a Data Space

1.1 Reference architecture model

The International Data Spaces Association (IDSA) offers a software architecture for Data Spaces with the abovementioned characteristics. IDSA created a Reference Architecture Model (RAM) (fig. 5) describing five levels and three perspectives.

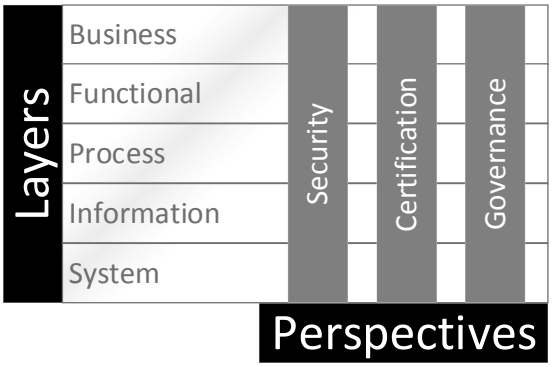

Fig. 5:
The five layers and three perspectives of the reference architecture model from International Data Spaces Association

Business Layer
In the business layer, the RAM defines the roles of the participants in the Data Space and categorizes them. It also specifies basic patterns of interaction between the roles and thus contributes to the development of innovative business models by the participants providing digital, data-driven services.
The business layer summarizes the roles into the four categories core participants, intermediaries, software developers, and Governance Bodies.

Functional Layer
The functional layer specifies the functional requirements for Data Spaces and the resulting functionalities to be implemented. The functional architecture is divided into six areas, each of which comprises a group of functionalities of the software.

Trust: Functional requirements must be met to achieve trust.
Security and Data Sovereignty: The functionalities regarding security and data sovereignty are fundamental in International Data Spaces.
*Ecosystem of Data: A Data Space needs d*ata source descriptions, meta data brokering and vocabularies.
Standardized Interoperability: This comprises functionalities for a standardized exchange of data between participants, which is the fundamental functionality in a Data Space.
Data Markets: To generate value, a Data Space must include concepts for clearing, billing and governance.

Information Layer
The information layer specifies the information model, the domain-independent, common language, i.e. the vocabulary used in the Data Space. The information model is an essential agreement that is shared by the participants and components of the IDS and facilitates compatibility and interoperability.

The main purpose of this formal model is to enable the (semi-)automatic exchange of digital resources within a trusted ecosystem of distributed parties while preserving the data sovereignty of the data owners. The information model supports the description, publication and identification of data products and reusable data processing software. Once the corresponding resources are identified, they can be exchanged and consumed via easily discoverable services. Apart from these core assets, the Information Model describes key elements of the International Data Space, its participants, its infrastructure components, and its processes.

Process Layer
The process layer specifies the interactions between the various components of the International Data Spaces. The most important processes are:

Onboarding: Onboarding describes what needs to be done to gain access to the Data Spaces
Data Offering: For the data offering a complete description of the data is an important prerequisite so that others can take advantage of the data.
Contract Negotiation: Connectors must negotiate the contract, i.e. accept data offers by negotiating the usage guidelines, which range from simple access restrictions to complex obligations before and after data exchange.
Exchanging Data: Connectors can only exchange data with each other once all previous process steps have been successfully completed. A connector calls up the data operation that relates to the contract agreement.
Publishing and Using Data Apps: Another relevant process is the publication and use of data apps. The apps are used to transform or process the data and are published by app providers in an app store and certified by the certification body if required. A connector can search for apps in the app store, download the appropriate app. The app is executed in the context of the connector.

System Layer
At the system level, the Data Space represents a distributed system in which the participating parties host the different Data Space components on their infrastructure. The connector initiates the exchange of data to and from the internal data resources of the participating organizations' enterprise systems and the Data Space. The connector provides metadata to the Metadata Broker as specified in the connector's self-description. This de-scription may include information about the technical interface, authentication mechanisms, and associated data usage policies. Usage agreements are transmitted by the connector to the Clearing House to establish trust. The Clearing House logs information about transactions to establish trust and transparency or to enable billing. The Vocabulary Hub provides vocabularies that specify the semantics of the exchanged data. Connectors can load applications from the App Store that run in the context of the connector and process data according to usage policies.

The System Layer maps the roles from the Business Layer and the processes from the Process Layer to a concrete data and service architecture. This represents the technical core of the Data Space (fig. 6).

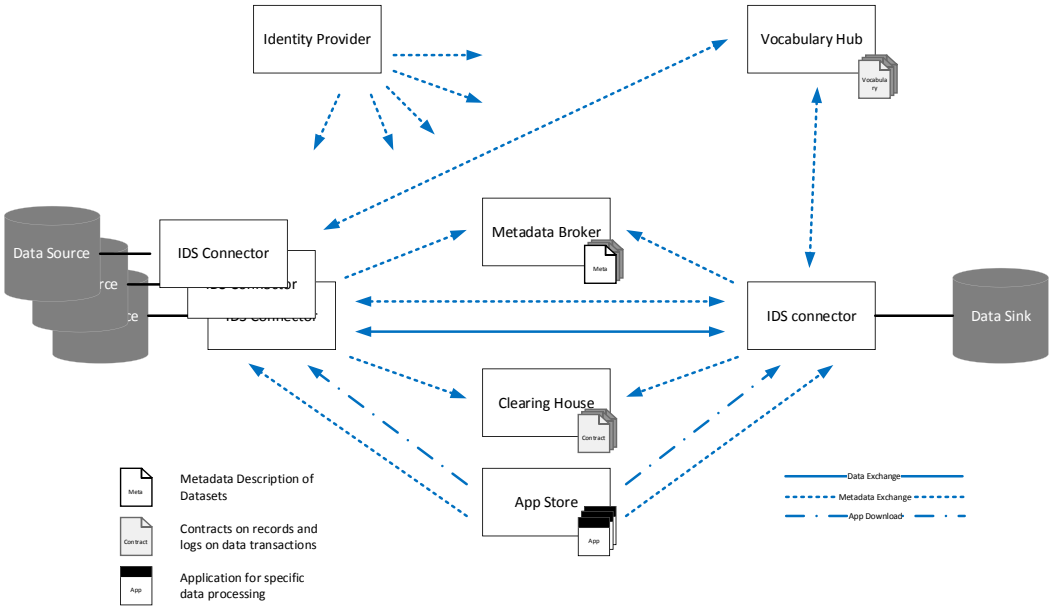

Fig. 6:
The system layer shows the interaction between the components of a Data Space and maps roles and processes to a data and service architecture.

Certification Perspective

To guarantee data security and data sovereignty, the Data Space relies on consistent certification of all components. In this way, the Data Space ensures that all players and components adhere to the rules of the Data Space. Certification applies to all levels of the architecture model.

The certification scheme outlines required levels and focuses for each role. At the functional level, core components must meet core security requirements for compliance certification. Conformity assessment includes the information level, functionality, protocols, and information model compatibility. Relevant processes are evaluated for compliance. At the system layer, certification focuses on security requirements, ensuring components are suitable for a trustworthy Data Space.

In addition to the Trust Levels, three Assurance Levels are applied as part of the certification process, ranging from self-assessment to external assessment.

To ensure the high quality and transparency of the certification process, all evaluation facilities must first be approved by the impartial certification body. This process comprises a compilation of all relevant information and documents and a check whether the evaluation facility complies with all the requirements of the Data Space and can be approved as an evaluation facility.

Security Perspective
A strategic requirement for Data Spaces is to provide secure data exchange. This is imperative to build and maintain trust between the components and organizations participating in the Data Space. A Data Space's security architecture must include measures to identify components in the Data Space, protect communication and data exchange, and control the use of the exchanged data. Connectors ensure that the specifications and requirements of the security architecture are applied to the interactions and operations in the Data Space.

Implementation of security depends on the layer, ranging from a security architecture on business level to technically implementing security specific hardware components on system level. For each level the measures to guarantee security must be adapted accordingly.

Which of the requirements are to be implemented depends on the required trust level. In the Data Space, a distinction is made between three trust levels: 1) interoperability in the Data Space, 2) complete function for controlling data usage, and 3) additional protection against internal attacks.

Governance Perspective
In the reference architecture model, the governance perspective defines the roles, functions and processes of the Data Spaces from the perspective of governance and compliance. A key point here is to describe the requirements to be fulfilled by the business ecosystem, which are necessary for secure and reliable interoperability between the companies. The architecture of Data Spaces does not impose any restrictions on cooperation between organizations and does not require compliance with predefined rules. The participating organizations determine the rules to be followed. The architecture as a functional framework enables the rules to be implemented.

The management of data-related resources through decision-making rights, responsibilities, roles, and ownership makes data governance a fundamental element in the Data Space ecosystem. Data governance makes it possible to successfully engage in a collaborative ecosystem. It is therefore important to create organizational structures and processes that clearly describe who can make what kind of decisions and what responsibilities are associated with these decisions.

From a governance perspective, data is an economic asset. Governance therefore focuses its activities on enabling new digital business models through the exchange of data. However, this considers the demands of data owners. This data ownership is another issue that must be respected by governance. In the legal sense, there is no ownership of data, as it is an intangible asset. Data sovereignty expresses the fact that the data owner always retains control over the use of their data.

2. Pooling Loads Using a Data Space

This chapter shows how aggregators can use a Data Space for the pooling of electrical loads to offer balancing energy on the market. In the context of Data Spaces, it is primarily the exchange of data that is of interest; the description does not go into the details of the provision of balancing energy.

The example describes the pooling of household appliances such as heat pumps, boilers, washing machines or charging stations for electric cars [4]. An aggregator requires data of a static (type, connected load, ...) or dynamic (status, availability, current power consumption, ...) nature for its service. To be able to provide the necessary power for balancing energy, many such devices with a low power consumption compared to classic balancing energy suppliers are required. The aggregator must be able to access this data to be able to offer the balancing energy at all. As the devices come from different manufacturers, the aggregator is faced with the challenge of being able to exchange data with devices from all manufacturers.

The solution approach using a Data Space envisages that each manufacturer's cloud is connected to the Data Space via a connector (fig. 7). These manufacturers are, in the sense of a Data Space, the data providers, providing the data coming from the devices. A service provider, such as the aggregator described above, is also connected to the Data Space via a connector. The aggregator is a data consumer. The onboarding procedure in the Data Space ensures the authenticity of the participating parties and provides a trustworthy framework in which the data can be exchanged by issuing certificates.

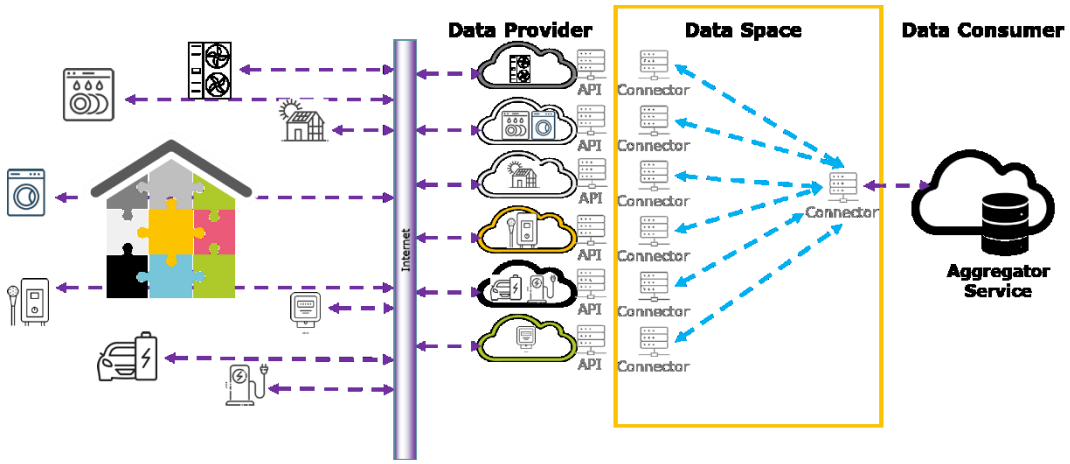

Fig. 7:
Solution approach for an aggregator using a Data Space.

The manufacturers of the devices (data providers) can make any of their data available in the Data Space, provided that the data owner declares his consent. To enable the aggregator (data consumer) to assess whether this data is useful for them, they can query the metadata provided by the manufacturers. In addition, the producers provide a description of their data (ontologies or vocabularies) so that the aggregator knows the meaning of the data. The aggregator then concludes a contract with the owners of the devices for the use of the data. In the case of pooling, the contract also stipulates that the owners also make their devices available as flexibilities. Such a contract not only includes that the aggregator may use the data, but the data owners also define how the aggregator may use the data using a data usage policy. These processes are shown in fig. 6.

Not all data has to be transferred to the aggregator. For example, it is conceivable that a consumption forecast for individual appliances could be calculated close to the manufacturer. In Data Spaces, an app running in the context of the connector can perform these calculations. This approach has the advantage that the aggregator does not need to receive the raw data in the first place and can therefore minimise the amount of data to be transmitted. The advantage for data owners is that the raw data does not even leave a controllable space, the Data Space, so privacy can be preserved as much as possible. The app store in the Data Space provides the apps.

The clearing house registers all transactions in the Data Space to be able to settle fees or costs. The information about the transactions only store what type of data was transferred and when, but not the data itself. The data owners thus always retain full transparency as to who requested which data from them and when. For the aggregator, the transactions serve to compensate the owners of the devices according to the use of the electrical loads.

To be able to use the devices in the pool for balancing energy applications, the aggregator must also be able to control them. A Data Space makes it possible not only to retrieve data from the manufacturer's clouds, but also to write data back to them. In this way, the aggregator can send information to the devices, which informs the devices about grid-friendly operation. It should remain under the control of the devices as to how they implement the information. In other words, whether they interrupt or resume operation immediately or whether they first complete a process that is currently running. The aggregator can then retrieve data on the current operating status via the Data Space.

3. Realization of a Data Space and Challenges

The authors have implemented a Data Space as a demonstrator as part of a project for the Swiss Federal Office of Energy (SFOE). They evaluated suitable software for this purpose. The International Data Spaces Association (IDSA) provides well-developed software with good documentation as open-source software. With the software components of the IDSA, it is possible to set up a Data Space in such a way that it fulfils all the requirements of the reference architecture model and thus represents a trustworthy space for exchanging data.

When implementing a Data Space, it has become clear that an extremely important requirement is the correct handling of certificates for all components involved in the Data Space. The certificates must be signed by a recognised, trustworthy authority. This is because only components certified in this way contribute to trustworthiness.

The process for exchanging data is highly complex, which leads to high implementation costs. The process consists of several steps from the negotiation of the data offer to the exchange of data, whereby security and sovereignty must be always guaranteed. The project team plans to provide a reference implementation so that the connection of data sources to the Data Space can be realized with less effort.

Challenges in the implementation of Data Spaces can also be expected at an organisational level. For example, ontologies are often difficult to understand and are not always available in a digitally processable form. This is the case if the ontology is available as a document from which the structure of the data model cannot be read automatically.

Although Data Spaces realize their full potential if the aggregated data sources can be used for many applications, many European projects are currently creating their own Data Spaces [5]. However, the European Data Strategy recommends realizing a Data Space for different domains (e.g. energy or mobility) [6]. However, these so-called Common Data Spaces are not yet sufficiently developed and do not represent a viable alternative in the projects.

4. Benefits and Potentials of Data Spaces

As already mentioned in the previous section, Data Spaces unfold their potential above all when a Data Space can be used for multiple use cases. A Data Space with many connected clouds represents a valuable source of data that service providers can use for their products. A Data Space can also make data available for research projects.

A major benefit of using a Data Space is that new business models can be realised without having to set up a complex infrastructure [7]. This requires the existence of a Data Space that provides the data required for the business model. If this condition is met, the service provider only needs to create a connector and have it certified. This represents a one-off effort, which is comparatively low when compared with the alternative of installing gateways at the customer's premises.

The concept of Data Spaces is based on trustworthiness by design. On the one hand, this can lead to many manufacturers participating in a Data Space as data providers. On the other hand, trustworthiness should also lead to data owners making their data available for the provision of services. In addition, the owners of the data retain their data sovereignty and can determine who is authorised to use the data and for what purposes.

5. Conclusion

The concept of Data Spaces is one way in which future digital applications can exchange data. Data Spaces are designed from the outset to be trustworthy, secure and sovereign. Data Spaces utilise an existing infrastructure that device manufacturers have already set up for data exchange. Data exchange does not result in data being stored multiple times. It remains at the data source and is only exchanged on request when required.

Certificates for participants in the Data Space and for components ensure trustworthiness. Authentication and encryption ensure security, while data usage policies together with the clearing house guarantee sovereignty.

Several challenges still need to be overcome before Data Space technology can be widely used. For example, it must be possible to reduce the effort required to implement a connector and connect to proprietary data sources. This can be done, for example, with a reference implementation that is made available to potential participants in the Data Space. For a smooth exchange of data, data providers and data consumers must agree on ontologies that are to apply in the Data Space. However, this requires digitally analysable representations of the ontologies, which currently applies in only a few cases.

With Data Spaces, new business models can be implemented in a new data economy without complex installations. The availability of a variety of different data sources enables the development of innovative or even disruptive business ideas. This potential must be utilised through the consistent further development and dissemination of Data Spaces.

References

[1] „Solar Energy“. Accessed 10. Mai 2024. https://energy.ec.europa.eu/topics/renewable-energy/solar-energy_en.

[2] BFE, Bundesamt für Energie. „Elektrizitätsproduktionsanlagen in der Schweiz“. Accessed 10. Mai 2024. https://www.uvek-gis.admin.ch/BFE/storymaps/EE_Elektrizitaetsproduktionsanlagen/.

[3] Otto, Boris, Michael ten Hompel, und Stefan Wrobel, Hrsg. Designing Data Spaces: The Ecosystem Approach to Competitive Advantage. Cham: Springer International Publishing, 2022. https://doi.org/10.1007/978-3-030-93975-5.

[4] „Equigy – Crowd Balancing Platform“. Accessed 21. Mai 2024. https://www.swissgrid.ch/de/home/newsroom/dossiers/crowd-balancing-platform.html.

[5] International Data Spaces. „Data Spaces Radar“. Accessed 21. Mai 2024. https://internationaldataspaces.org/adopt/data-spaces-radar/.

[6] „European Data Strategy - European Commission“. Accessed 21. Mai 2024. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en.

[7] International Data Spaces. „Data Spaces“. Accessed 21. Mai 2024. https://internationaldataspaces.org/why/data-spaces/.