World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

Data sovereignty frameworks for space-based data platforms

Vaghani Divyeshkumar *

Gannon University, 109 University Square, Erie, PA 16541, USA.

World Journal of Advanced Research and Reviews, 2024, 22(03), 653–664

Publication history: Received on 01 May 2024; revised on 09 June 2024; accepted on 11 June 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.22.3.1770

Abstract

Despite widespread support, data sovereignty is still in its nascent stages. Consequently, current methods for transferring sovereign data are underdeveloped and could be enhanced. To address these issues, this paper introduces an architecture that utilizes edge computing resources to host both the data and the connector. Additionally, we develop a system based on this architecture, enabling the transfer of sovereign data between users using space-based data platforms. To assess our approach, we deploy our system in a real-world environment with users located in different countries and conduct various experiments. In order to generate representative outcomes, we conduct a total of 2,520 experiments. The results demonstrate significant performance improvements, such as reduced communication latency and increased network bandwidth, compared to a baseline utilizing cloud computing resources. We have constructed a system for sovereign data exchange that leverages edge computing resources to facilitate the transfer of data from a provider to a consumer. Furthermore, we have deployed this system in a real-world scenario and have demonstrated that the proposed approach offers reduced latency and increased bandwidth compared to a baseline relying on the cloud. Given the promising outcomes, we believe that future investigations could delve into usage control techniques that can be implemented at the network edge and explore scalability considerations.

Keywords: Cloud Computing; Data Sovereignty; Edge Computing; Scalability; Space-Based Platforms.

1. Introduction

Data Sovereignty is an emerging concept in the distributed systems field, concerning the distribution and usage of data [1]. According to this concept, data is governed by the laws of the nation where it is collected, and the data provider sets specific usage constraints, including who can use it, in what context, and how [1]. This idea is gaining traction as governments recognize that, due to cloud computing, data collected locally (e.g., from citizens) may be processed abroad, potentially ignoring the privacy regulations of the source country [2], [3]. Additionally, data stored outside its country of origin becomes subject to the host country's laws, which might lead to the exposure of sensitive information to that government [2]. These concerns drive the creation of protective legislation like the European GDPR [4] and the USA's ban on Huawei [1]. Moreover, nations (e.g., Canada) issue white papers and guidelines to enhance awareness of data sovereignty and safeguard citizen privacy [5].

To adhere to data sovereignty principles, various initiatives involving both academic and industry partners have emerged [1], [6], [7]. Their main objective is to develop architectures and mechanisms that support the transfer of sovereign data, meaning data that retains its sovereignty during transfer [1]. Initiatives such as the International Data Spaces Association (IDSA) and Gaia-X promote the adoption of sovereign data through software, documentation, and events [8].

Despite widespread support, data sovereignty is still in its nascent stages. Consequently, current methods for transferring sovereign data are underdeveloped and could be enhanced. For instance, in existing architectures, data is

* Corresponding author: Vaghani Divyeshkumar

typically not stored in the cloud due to sovereignty issues. However, other components, like authentication mechanisms or system monitoring, may be cloud-based [9], [10], [11]. Another potential cloud component is the connector that facilitates data transfer [12]. While placing components in the cloud can offer benefits (as discussed in Section IV-A), it may also lead to high latency and network bandwidth issues [13]. These drawbacks can negatively impact user experience and impede the adoption of data sovereignty systems.

To address these issues, this paper introduces an architecture that utilizes edge computing resources to host both the data and the connector. Additionally, we develop a system based on this architecture, enabling the transfer of sovereign data between users without relying on cloud services. To assess our approach, we deploy our system in a real-world environment with users located in different countries and conduct various experiments. The results demonstrate significant performance improvements, such as reduced communication latency and increased network bandwidth, compared to a baseline utilizing cloud computing resources.

2. Literature Review

Research on data sovereignty is covered in conceptual papers, reviews, white papers, and reports [10], [14]. Geisler et al. [15] offer a detailed overview of data ecosystems, examining the requirements and challenges of data handling with a focus on privacy and sovereignty. Kotka et al. [16] explore the legal issues of transferring sensitive data to the cloud, considering data sovereignty and protection aspects. Firdausy et al. [12] address data sovereignty's applicability to enterprises and organizations. Solmaz et al. [17] discuss data spaces—controlled environments for exchanging sovereign data—highlighting motivation, technical developments, and interoperability challenges. Brost et al. [18] emphasize the value of data in industrial data spaces, focusing on security and usage control techniques to ensure data use aligns with provider stipulations. While these studies contribute significantly to data sovereignty, they lack quantitative results from real-world implementations, which this work aims to provide.

There is also related research involving implementations and results, though it is less extensive. Qarawlus et al. [19] and Nast et al. [20] address the challenge of managing sovereign data on devices with limited hardware capabilities. Qarawlus et al. focus on messaging schemes, whereas Nast et al. design a specialized connector exposing sovereign data via a standardized API. Sarabia-Jacome et al. [21] present a system for sovereign data exchange targeting a seaport scenario, where data from the port terminal is shared with the port authority using the FIWARE platform [22]. Liang et al. [23] propose a system for sharing personal health data, considering privacy and sovereignty. In their system, health data from wearable devices is uploaded to the cloud, and access requires explicit user consent, such as from an insurance company. These approaches, however, do not discuss the impact of placing the connector in the cloud versus at the edge. To our knowledge, this paper is one of the first to provide empirical results on the latency of transferring sovereign data using edge resources.

2.1. Data Sovereignty

In 2015, the "Digital Sovereignty in a Connected Economy" Focus Group introduced the idea of "data sovereignty." At that time, as now, "the concept of sovereignty should be initially understood in broad terms as the capacity for self-determination, manifesting as independence and autonomy. In this context, digital sovereignty refers to the ability to determine actions and make decisions within the digital realm." [45].

Since then, the concept has been further elaborated and categorized as a facet of general sovereignty. Consequently, the management of data and core technological capabilities became a crucial topic at the 2018 Digital Summit: "The data sovereignty of a state or organization inherently involves complete control over stored and processed data, and autonomous decision-making regarding who can access it." [46]. This also encompasses the "ability to independently develop, modify, and control technical components and systems, and to integrate them with other components." [46]. The aim of digital sovereignty is not absolute but allows for varying degrees. A stratified model was created based on the criteria identified in 2018.

Data sovereignty is essential for the independent operation of both the state and the commercial sector. It enhances economic sustainability, competitiveness, agility, and risk management capabilities. States and organizations that possess digital sovereignty can operate more freely in the market, being less reliant on manufacturers or suppliers. Additionally, due to lower barriers to market entry, data sovereignty enables companies and organizations to more effectively act as suppliers within digital ecosystems, thereby gaining opportunities to innovate and influence developments. Overall, systems, processes, and interactions can be more readily adapted, developed, and, if necessary, replaced.

2.2. Platform-Based Ecosystems

The idea of "platform-based ecosystems" stems from the analogy of natural ecosystems, which serve as habitats for various organisms and their surroundings. Platform-based ecosystems are thus "defined as identifiable value networks built upon an existing [technical] network architecture" [47].

Platforms are often developed and maintained to establish new business models, especially in the B2C sector, where organizations are driven by the desire to implement associated business models. Here, the motivation, creation, and management are handled by the same entity.

In contrast, in the B2B sector, platforms might be established to ensure secure communication or the secure transfer of data and goods. These platforms can be initiated by industrial consortia, associations, or cooperatives, with one company being tasked with their creation and operation, thus acting as the technical platform provider. It is, therefore, useful to differentiate between technical and economic platform operators.

Economic platform operators, including public-sector operators, are those whose primary interest lies in the platform's existence and who have a vested economic interest in defining the rules for interaction and the platform's technical design. Technical operators, on the other hand, provide the necessary technical framework, implementation, and seamless operation of the IT infrastructure and interfaces for the platform. Together, these entities form the core of platform-based ecosystems. Other common actors (periphery) include providers of content, goods, or applications, and the users. Interactions within and between digital ecosystems are facilitated by interfaces and standards where needed (see Fig. 1).

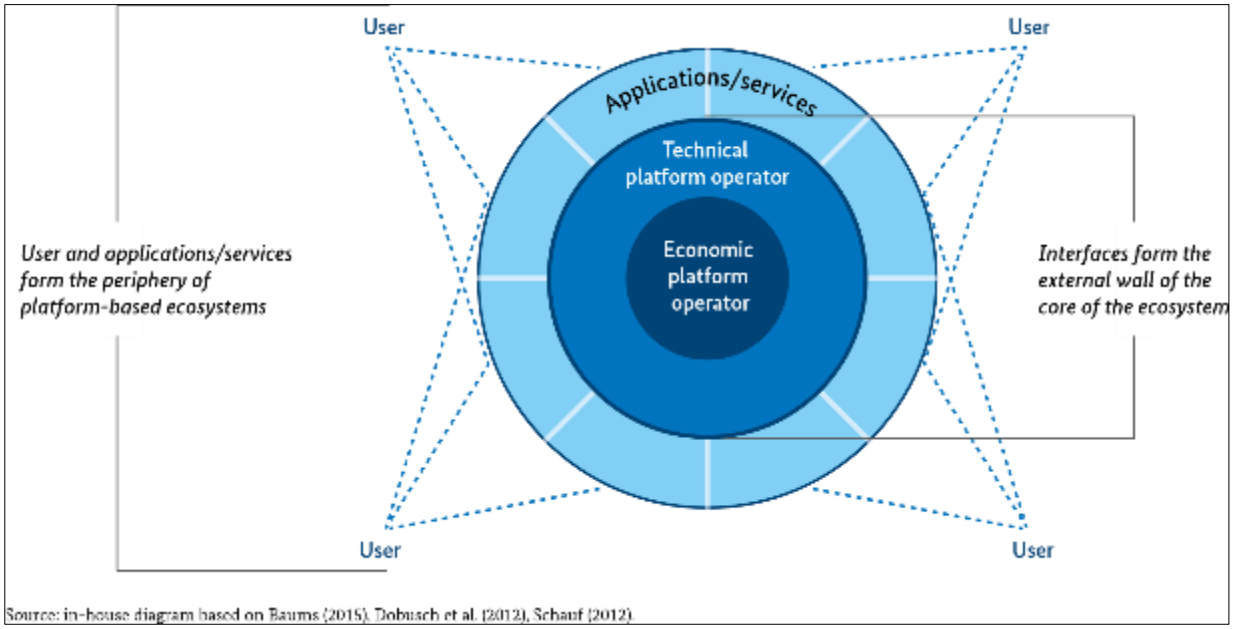

Figure 1 Schematic depiction of a platform-based digital ecosystem

A unifying feature of all platform-based ecosystems is their reliance on digital platforms. These platforms support the secure and trusted execution of transactions to offer, obtain, or use goods, services, content, information, and data between people and/or technical systems.

2.3. System Model

In this section, we outline a system model designed to facilitate the exchange of sovereign data. Our model is based on well-established reference architectures from the literature and adheres to the principles of data sovereignty [6], [10], [24]. To enhance clarity, we first introduce the terminology. Then, we describe the key components of the system and the basic interactions among them.

2.3.1. Nomenclature

In a system for sovereign data exchange, there are data consumers and data providers. A consumer is an entity, such as an individual, organization, or application, that seeks to acquire sovereign data and use it in accordance with its terms of use. A provider is an individual or organization that owns data and wishes to share it with others, under specific terms of use. Sovereign data, or a sovereign dataset, is a dataset accompanied by terms of use established by the data provider, compliant with local laws and regulations. These terms dictate the permissible uses of the data, specifying the types of allowed processing (e.g., statistical analysis, machine learning), the duration of such processing (e.g., one month), redistribution rules, and other constraints. The complete set of these terms is known as the dataset's policy. Policies can be formulated using ODRL (Open Digital Rights Language), which provides a framework for expressing rights in JSON or XML.

To manage the acquisition and acceptance of policies, a connector can be employed [25]. The connector is a software application that facilitates the basic interactions required for exchanging sovereign data, such as requesting a dataset, agreeing to the policy, and transferring the dataset from the provider to the consumer. Thus, the connector also helps achieve interoperability between data providers and consumers.

Beyond the connector, a system for sovereign data exchange may include federated services provided by a trusted organization [24]. These federated services are software applications available to all participants, including both providers and consumers. Two key federated services are the trusted authority and the clearing house [10]. The trusted authority verifies the identity of connectors and ensures trust between providers and consumers. The clearing house records all interactions within the system and helps resolve disputes (e.g., if a consumer claims a dataset was not transferred, but the provider asserts it was). These two federated services aim to ensure that for any sovereign data transfer, there are records of the participants' identities and the agreed policy.

2.3.2. Architecture

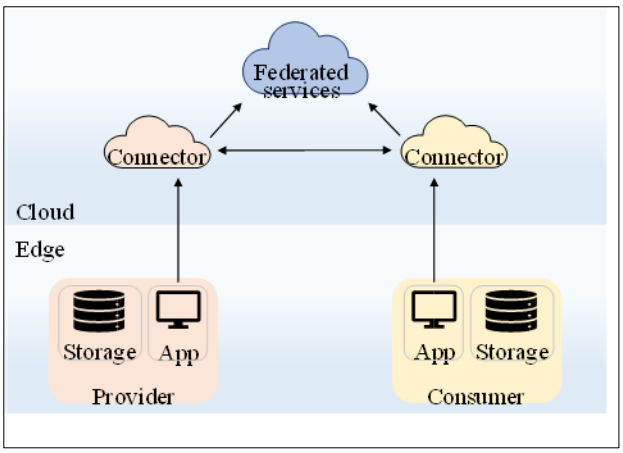

Figure 2 System architecture for sovereign data exchange.

A system that enables the exchange of sovereign data comprises various components, which can be deployed either in the cloud (e.g., using commercial providers like Google) or at the edge (e.g., using private on-premise computing resources), as illustrated in Fig. 2. Although such a system might include multiple data providers and consumers, Fig. 2 provides a simplified representation with a single provider and a single consumer. In this setup, the provider stores sovereign data in local storage. Similarly, the consumer has local storage for hosting the acquired data. Both the provider and the consumer have a local application (i.e., a front-end application) that offers a user interface to the system. Through this app, the provider can interact with the provider connector, and the consumer can interact with the consumer connector. Typically, both connectors are deployed in the cloud (this will be further discussed in Section IV-A). The federated services, intended for use by all participants, are also cloud-based [10], [24].

To initiate the exchange of sovereign data, the provider first registers a dataset with the provider connector. This involves submitting the dataset's policy and the means to retrieve the dataset from the provider’s storage (e.g., via a protocol such as HTTP). Once the dataset is registered with the provider connector, consumers can request it. The consumer (using the consumer app) sends a request to the consumer connector to obtain a dataset registered with the provider connector [6]. The steps involved in requesting and transferring a sovereign dataset are outlined in Fig. 3. In Step 1 of Fig. 3, the consumer connector requests a dataset from the provider connector. The provider connector then

responds with the dataset's policy (Step 2). The consumer reviews this policy and signs a contract agreeing to the terms of use (Step 3). The provider connector acknowledges this with a contract agreement (Step 4). The consumer connector subsequently requests the transfer of the agreed dataset (Step 5), providing a file path to the consumer's storage (and any necessary credentials). Finally, the provider connector retrieves the dataset from the provider's storage (Steps 6 and 7) and transfers it to the consumer's storage (Steps 8 and 9), completing the transfer (Step 10).

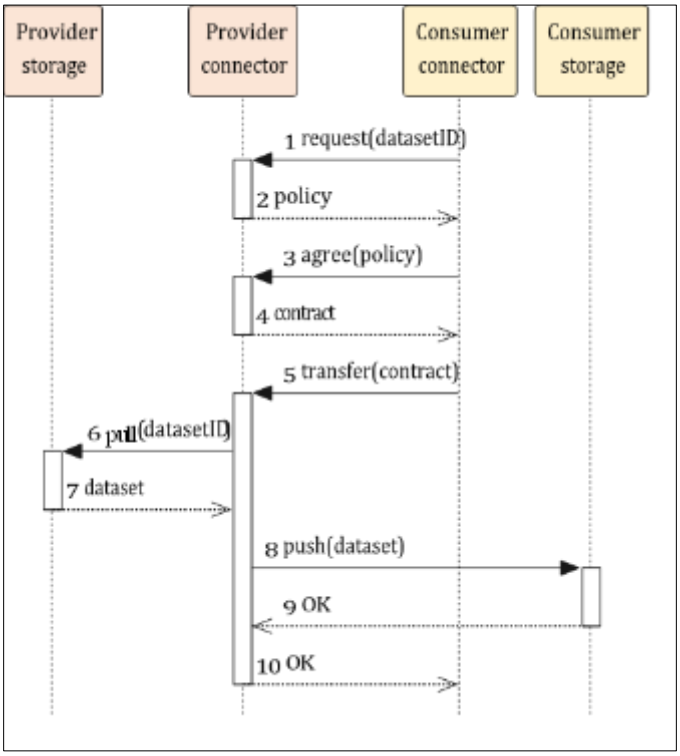

Figure 3 Transfer of a sovereign dataset.

While Fig. 3 illustrates interactions among the core components of the system, additional interactions with federated services may be required. For instance, a consumer may need to search for relevant sovereign datasets offered by various providers. These searches can be conducted through a federated service acting as a central catalog where providers advertise their data (e.g., using metadata or self-descriptions [24]). By browsing this catalog, the consumer can locate the addresses of provider connectors with useful datasets and request these datasets through the consumer connector (as depicted in Fig. 3).

2.4. Sovereign Data with Edge Computing

In this section, we propose an architecture for enabling sovereign data that leverages edge resources to enhance functionality and performance. The subsequent sections presents key observations regarding the placement of system components and the potential benefits of using edge computing resources, and also describes the proposed architecture.

2.4.1. Observations

As depicted in Fig. 3, both the provider and the consumer utilize on-premise storage for hosting sovereign datasets. This approach is crucial because using third-party storage (e.g., cloud services) can undermine data sovereignty principles [3]. For instance, a sovereign dataset must be used in accordance with its policy. However, storage providers typically do not offer policy guarantees. Instead, users of the storage must usually comply with the storage provider's terms of use, which may include anonymized data processing without considering each dataset's policy [3]. Therefore, on-premise storage is preferred for sovereign data.

Connectors, however, can be deployed either in the cloud or at the edge [12], [26]. Despite this flexibility, early implementations tend to focus on cloud deployments [27], [28]. There are several reasons for this trend. The cloud offers scalable, general-purpose resources that are cost-effective and eliminate the need to purchase or maintain equipment [29]. Consequently, using the cloud for connectors provides ease, flexibility, and scalability [30]. Connectors store information about datasets and policies but do not store the actual data. During data transfer, the provider

connector streams data from the provider storage to the consumer storage without storing it. Thus, deploying connectors in the cloud does not violate sovereignty principles, provided the cloud's location is not restricted.

Interestingly, the ability to run connectors in the cloud has opened new business opportunities for platform-as-a-service models [28], also known as Data Sovereignty-as-a-Service or Connector-as-a-Service [31]. In these models, the service provider hosts a connector instance in the cloud and manages all necessary maintenance and interactions with federated services. By using this service, users can fully utilize a sovereign data exchange system without needing to run and maintain a connector themselves.

Instead, users only need to register datasets and define policies.

While deploying connectors in the cloud can be justified for various valid reasons, it can also have some drawbacks. When the provider connector retrieves data from the source (i.e., the provider's storage) and transfers it to the destination (i.e., the consumer's storage), there is no consideration for the network path, transfer latency, or bandwidth utilization. This can lead to situations where the source and destination are physically close, but the data still travels through a distant cloud, consuming additional network resources and increasing transfer latency [32]. Although this specific issue has not been extensively studied in the context of data sovereignty, similar challenges have been addressed in edge computing and IoT contexts [33]-[38]. Therefore, in the next section, we propose an architecture for sovereign data exchange inspired by edge computing and highlight the potential benefits.

2.4.2. Proposed System Architecture

To formulate the proposed system architecture, we begin by examining a typical scenario of sovereign data transfer, as illustrated in Fig. 3. During the transfer process, when the provider connector retrieves data from the provider's storage and transfers it to the consumer's storage, the communication latency comprises two components: the latency from the provider to the cloud (LatPro→Cloud) and the latency from the cloud to the consumer (LatCloud→Con). Consequently, the overall transfer latency through the cloud, denoted as LatCloudPro→Con, can be expressed as:

$$LatCloudPro \rightarrow Con = LatPro \rightarrow Cloud + LatCloud \rightarrow Con \quad (1)$$

Thus, the cloud essentially serves as a detour on the network path between the provider and the consumer. Detours through the cloud often result in significant consumption of network resources and increased latency [32]. For this reason, we propose relocating the connectors to on-premise locations, as depicted in Fig. 4. In this setup, the provider connector and the consumer connector are situated at the provider's and consumer's sites, respectively. Despite this change, users can still access the connectors through their respective applications, and the connectors can maintain communication with federated services deployed in the cloud. However, with both connectors positioned at the network's edge, they may be able to communicate directly with each other without traversing a remote cloud. This has the potential to yield performance enhancements and additional functionality associated with the utilization of edge resources (as elaborated below).

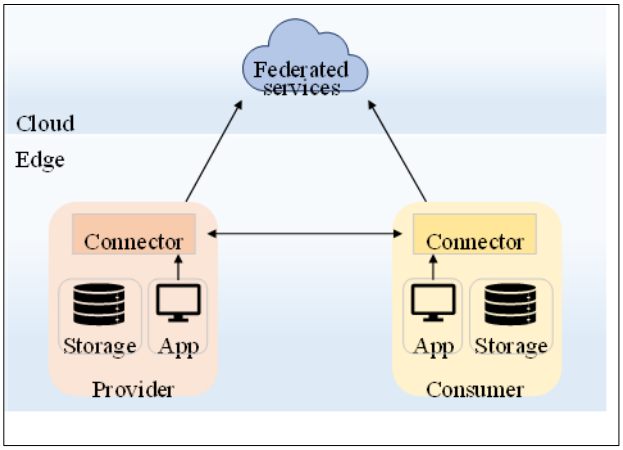

Figure 4 System architecture with on-premise connectors.

In terms of functionality, with direct data transfer from the provider to the consumer in the proposed architecture, there is no risk of routing data through a cloud situated in an unauthorized region. For instance, if both the provider and the

consumer are located in Austria and there's a requirement for data not to leave the country, running the provider connector in a cloud located in another country (e.g., Switzerland) would violate this requirement as the data traverses through the provider connector's cloud location. Such issues are circumvented in the proposed architecture, where the provider connector is placed on the provider's premises, enabling direct data transmission from the provider site to the consumer site, bypassing the cloud.

Regarding performance, the communication latency of the transfer in the proposed architecture encompasses the latency of the provider connector to retrieve data from storage and send it to the consumer (LatPro→Con). When the provider connector is at the provider site, the former latency can be considered negligible since there's no communication latency, only reading a file from the filesystem. Hence, the transfer latency when the connectors are at the edge, denoted as LatEdgePro→Con, is:

$$LatEdgePro \rightarrow Con = 0 + LatPro \rightarrow Con \quad (2)$$

Comparing equation 1 with equation 2, it's observed that both latencies involve the same source and destination. However, equation 1 involves a detour through the cloud. If the cloud lies on the path from source to destination, the network path in equation 1 will be similar to that in equation 2 [39]. Consequently, the two latencies will be comparable, i.e., $LatEdgePro \rightarrow Con \approx LatCloudPro \rightarrow Con$ (3). However, if the cloud is not on the path between the source and destination, the network path in equation 1 will be longer (due to the detour), resulting in higher latency [32].

Hence, $LatEdgePro \rightarrow Con < LatCloudPro \rightarrow Con$ (4). By combining equations 3 and 4, it can be inferred that:

$$LatEdgePro \rightarrow Con \lesssim LatCloudPro \rightarrow Con \quad (5)$$

Thus, the proposed architecture is anticipated to offer similar or lower communication latency for transferring sovereign data. Potentially, utilizing edge computing resources may also yield other benefits, such as enhanced user experience due to reduced response times when users interact with the connector (via the app) since the connector is now deployed closer to the app for both the provider and the consumer, as illustrated in Fig. 3. Nonetheless, in this paper, and also in our evaluation in Section V, the focus remains on benefits related to data transfer, which is the primary objective of data sovereignty.

2.5. Evaluation

To assess our methodology, we construct a system adhering to the proposed edge architecture (depicted in Fig. 4). Additionally, we develop a system following the conventional cloud architecture (illustrated in Fig. 3) for use as a reference point. Our evaluation aims to compare the two approaches in terms of communication latency and network bandwidth. For this purpose, we deploy connectors in the cloud using the Google Cloud Platform, which serves as the baseline. Furthermore, we deploy connectors at user sites, situated in various countries, as per our proposed approach. These user sites incorporate a custom Python-based storage application for data hosting. Fig. 5 depicts the cloud's location in Germany and user sites spread across Spain, England, France, Belgium, Netherlands, Italy, and Poland, all within Europe and varying distances from the cloud.

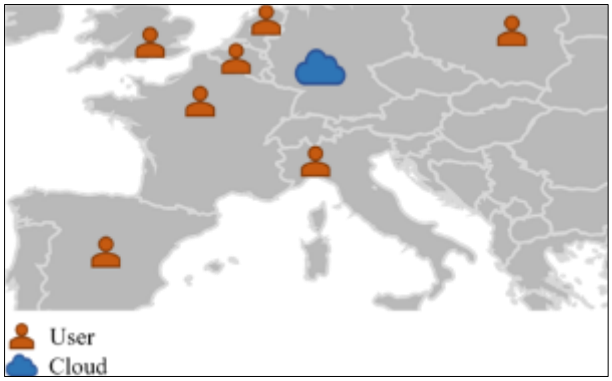

Figure 5 Location of the cloud and the users in Europe.

We deem this geographical area suitable for our experiments due to the existing involvement of many European countries in testing sovereign data methodologies. The connectors utilized are based on the Eclipse Dataspace

Connector, an open-source Java-based software supported by the Eclipse foundation. We make certain modifications to ensure compatibility with our system. All components are interconnected via the Internet.

For this evaluation, we select a smart energy use case, as smart energy applications typically involve user interactions, such as alerts or real-time analytics, and rely on low latency to provide prompt response times and enhance user experience. Given that achieving low latency is our primary objective, we deem this use case appropriate. We utilize data from a publicly available dataset containing real measurements from smart meters. Various data sizes are considered to represent different application scenarios: 85 Bytes (B) for a single measurement, 7 Kilobytes (KB) for daily measurements, 224 KB for monthly measurements, 1 Megabyte (MB) for 4.5 months, 2 MB for 9 months, and 4 MB for 18 months. For instance, the data size of a single measurement could represent a safety application wherein a household (provider) sends sovereign data to an anomaly-detection service (consumer) for real-time detection of potential hazards like gas leaks or malfunctions. Although smaller data sizes may seem more relevant for such cases to transmit new measurements promptly, applications utilizing larger data sizes for real-time analytics also require low latency. Hence, larger data sizes are also considered.

2.5.1. Findings

In order to generate representative outcomes, we conduct a total of 2,520 experiments. In the baseline experiments, the provider connector, situated in the cloud, follows the process outlined in Fig. 3 to transfer data from the provider's site to the consumer's site via the cloud. Conversely, in the proposed approach experiments, the provider connector, now located at the provider's site, undergoes the same process to directly transfer data from the provider's site to the consumer's site. In both approaches, each user serves as a data provider, transmitting sovereign data of varying sizes to all other users (acting as consumers in different countries) multiple times. For each transmission, we measure the communication latency and the hop count of the transfer.

To illustrate the latency measurements, we present Table I and Fig. 6. Fig. 6 visually represents the latency for each data size under both the baseline and proposed approaches. Meanwhile, Table I provides the numerical values of the average latency, standard deviation, and the reduction in the average achieved by the proposed approach. As depicted in Fig. 6, the communication latency for each approach is comparable for very small data sizes and increases with larger data sizes. Our proposed approach exhibits a similar standard deviation to the baseline, indicating a similar distribution of values. However, the average latency in our approach is approximately 20% lower than the baseline for all data sizes (also detailed in Table I). Additionally, it is noteworthy that the upper quartile of our proposed approach consistently equals or falls below the baseline's average. This implies that in our approach, 75% of the values are either similar to or lower than the average value of the baseline. Furthermore, the minimum values for every data size in our proposed approach, represented by the lower whiskers in Fig. 6, are consistently at least 50% lower than the corresponding minimum values in the baseline.

Table I Communication latency in milliseconds (ms) of each data size and the percentages of reduction**.**

	Baseline		**Proposed**		**Reduction**
	avg	st dev	avg	st dev	avg (%)
85 B	63	15	51	16	19
7 KB	63	17	51	17	19
224 KB	173	51	133	52	23
1 MB	249	73	198	74	20
2 MB	266	79	219	82	18
4 MB	313	84	258	89	18

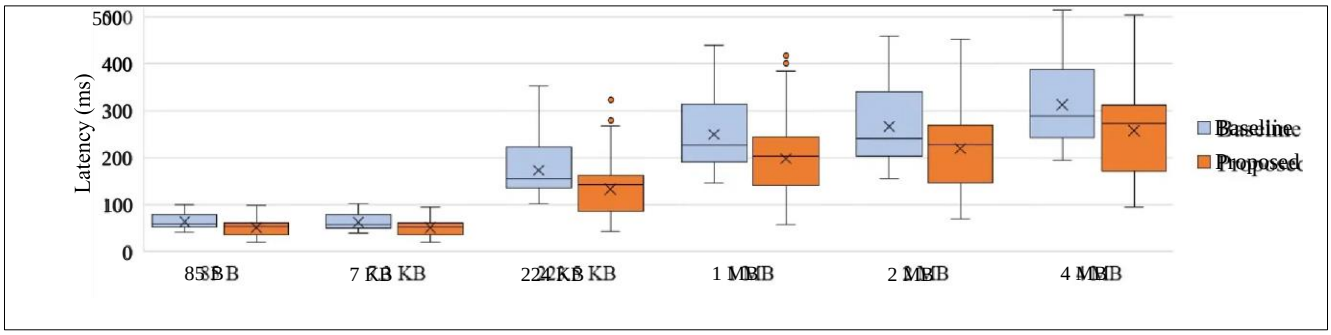

Figure 6 Communication latency (in ms) of each data size based on the baseline and the proposed approach**.**

Remarkably, such a notable decrease in latency (i.e., 50% or more) occurs for data transfers between closely located users, such as those in Belgium and the Netherlands, due to the significant detour created when transferring data through the cloud. It is also observed that the maximum latency values for all data sizes remain consistent across both approaches. This is evident when the network path in both approaches is similar, particularly when the cloud is on the path from the provider to the consumer. In our experiments, the highest latency is recorded for transfers between users in Spain and Poland. In this scenario, transferring data through the cloud does not result in a noticeable detour, as illustrated in Fig. 5. Lastly, it is noted that while the percentage reduction in average latency remains relatively stable (around 20% as indicated in Table I), the actual reduction in milliseconds increases with larger data sizes. The average latency reduction begins at 12 milliseconds for data of 85 bytes and escalates to 55 milliseconds for data of 4 megabytes. Given that the examined use case aims to detect potential hazards in real time, a 20% latency reduction, potentially reaching 50%, can be deemed a substantial enhancement. Overall, the results align with the latency analysis, demonstrating that the proposed edge architecture tends to diminish communication latency compared to the cloud architecture.

To further explore the transfer of sovereign data, we assess the hop count of data transfers in both methods, as a high number of hops can indicate delay and bandwidth limitations [13], [44]. Consequently, transfers over shorter network paths (i.e., with fewer hops) are anticipated to have more available bandwidth. In our experiments, the number of hops between a provider and a consumer remains consistent across all data sizes, thus we do not individually plot the hops for each data size. Fig. 7 illustrates the hop count for the two approaches under examination. The baseline exhibits an average of 31 hops (with a standard deviation of 2), while the proposed approach averages 18 hops (with a standard deviation of 5). This results in a 42% decrease in the average, with all values of the proposed approach being lower than those of the baseline, except for outliers. Therefore, overall, the proposed approach utilizes shorter paths than the baseline, indicating that the transfer of sovereign data using our approach is likely to have more available bandwidth.

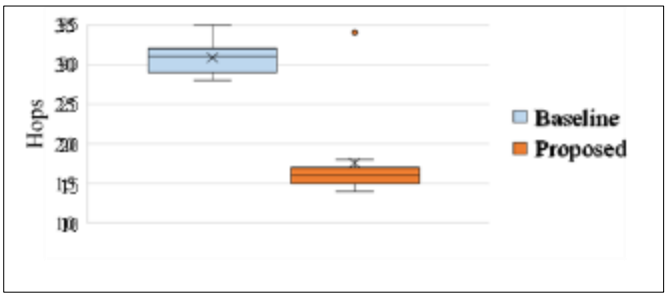

Figure 7 Hop count of the data transfers based on the baseline and the proposed approach.

Regarding computational resource utilization, it is observed that both approaches perform similarly; for instance, CPU utilization in both cases remains below 5% with occasional spikes. This is because the same number of data transfers is executed in both scenarios. The primary distinction between the two methods is that the baseline necessitates the deployment of two connectors (a provider connector and a consumer connector) in the cloud to serve all users. Conversely, the proposed approach mandates two connectors (a provider connector and a consumer connector) to be deployed on-premise for each user. Consequently, the former entails the continuous operation of connectors to serve all users, while the latter results in some connectors being idle when they are not engaged in data transfers.

3. Conclusion

In conclusion, we have constructed a system for sovereign data exchange that leverages edge computing resources to facilitate the transfer of data from a provider to a consumer. Furthermore, we have deployed this system in a real-world scenario and have demonstrated that the proposed approach offers reduced latency and increased bandwidth compared to a baseline relying on the cloud. Given the promising outcomes, we believe that future investigations could delve into usage control techniques that can be implemented at the network edge and explore scalability considerations.

References

[1] A. Braud, G. Fromentoux, B. Radier, and O. Le Grand, "The road to European digital sovereignty with Gaia-X and IDSA," *IEEE Network*, vol. 35, no. 2, pp. 4–5, 2021.

[2] "Government of Canada white paper: Data sovereignty and public cloud," in *https://www.canada.ca/en/government/system/digitalgovernment/digital-government-innovations/cloud-services/gc-whitepaper-data-sovereignty-public-cloud.html*. Accessed: January 2023.

[3] C. R. Baudoin, "The impact of data residency on cloud computing," in *Proceedings of the International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, pp. 430–435, IEEE, 2018.

[4] K. Singi, S. G. Choudhury, V. Kaulgud, R. J. C. Bose, S. Podder, and A. P. Burden, "Data sovereignty governance framework," in *Proceedings of the International Conference on Software Engineering Workshops (ICSE)*, pp. 303–306, ACM, 2020.

[5] M. Lukings and A. Habibi Lashkari, "Data sovereignty," in *Understanding Cybersecurity Law in Data Sovereignty and Digital Governance*, pp. 1–38, Springer, 2022.

[6] "IDS reference architecture model (version 3.0)," pp. 1–118, International Data Spaces Association, 2019.

[7] F. Lauf, S. Scheider, J. Bartsch, P. Herrmann, M. Radic, M. Rebbert,

[8] A. T. Nemat, C. Schlueter Langdon, R. Konrad, A. Sunyaev, and S. Meister, "Linking data sovereignty and data economy: arising areas of tension," *Wirtschaftsinformatik Proceedings 19*, 2022.

[9] B. Otto, "A federated infrastructure for European data spaces," *Communications of the ACM*, vol. 65, no. 4, pp. 44–45, 2022.

[10] T. Uslander, M. Baumann, S. Boschert, R. Rosen, O. Sauer, L. Sto-¨ janovic, and J. C. Wehrstedt, "Symbiotic evolution of digital twin systems and dataspaces," *Automation*, vol. 3, no. 3, pp. 378–399, 2022.

[11] "Position paper: GAIA-X and IDS," pp. 1–33, International Data Spaces Association, 2021.

[12] N. Naik and P. Jenkins, "Self-sovereign identity specifications: Govern your identity through your digital wallet using blockchain technology," in *Proceedings of the International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud)*, pp. 90–95, IEEE, 2020.

[13] D. R. Firdausy, P. D. A. Silva, M. Van Sinderen, and M.-E. Iacob, "Towards a reference enterprise architecture to enforce digital sovereignty in international data spaces," in *Proceedings of the Conference on Business Informatics (CBI)*, vol. 1, pp. 117–125, IEEE, 2022.

[14] M. Satyanarayanan, "How we created edge computing," *Nature Electronics*, vol. 2, no. 1, pp. 42–42, 2019.

[15] "White paper: Edge computing in the EuProGigant project," pp. 1–20, Institute for Production Management, Technology and Machine Tools (PTW), Technical University of Darmstadt, 2021.

[16] S. Geisler, M.-E. Vidal, C. Cappiello, B. F. Loscio, A. Gal, M. Jarke,´ M. Lenzerini, P. Missier, B. Otto, E. Paja, *et al.*, "Knowledge-driven data ecosystems toward data transparency," *ACM Journal of Data and Information Quality (JDIQ)*, vol. 14, no. 1, pp. 1–12, 2021.

[17] T. Kotka, L. Kask, K. Raudsepp, T. Storch, R. Radloff, and I. Liiv, "Policy and legal environment analysis for e-government services migration to the public cloud," in *Proceedings of the International Conference on Theory and Practice of Electronic Governance (ICEGOV)*, pp. 103–108, ACM, 2016.

[18] G. Solmaz, F. Cirillo, J. Furst, T. Jacobs, M. Bauer, E. Kovacs, J. R.¨ Santana, and L. Sanchez, "Enabling data spaces: existing developments´ and challenges," in *Proceedings of the International Workshop on Data Economy (DE)*, pp. 42–48, ACM, 2022.

[19] G. S. Brost, M. Huber, M. Weiß, M. Protsenko, J. Schutte, and S. Wessel,¨ "An ecosystem and IoT device architecture for building trust in the industrial data space," in *Proceedings of the Workshop on CyberPhysical System Security (CPSS)*, pp. 39–50, ACM, 2018.

[20] H. Qarawlus, M. Hellmeier, J. Pieperbeck, R. Quensel, S. Biehs, and M. Peschke, "Sovereign data exchange in cloud-connected IoT using international data spaces," in *Proceedings of the Cloud Summit*, pp. 13– 18, IEEE, 2021.

[21] M. Nast, B. Rother, F. Golatowski, D. Timmermann, J. Leveling, C. Olms, and C. Nissen, "Work-in-progress: Towards an international data spaces connector for the internet of things," in *Proceedings of the International Conference on Factory Communication Systems (WFCS)*, pp. 1–4, IEEE, 2020.

[22] D. Sarabia-Jacome, I. Lacalle, C. E. Palau, and M. Esteve, "Enabling´ industrial data space architecture for seaport scenario," in *Proceedings of the World Forum on Internet of Things (WF-IoT)*, pp. 101–106, IEEE, 2019.

[23] V. Araujo, K. Mitra, S. Saguna, and C. Ahlund, "Performance evaluation˚ of FIWARE: A cloud-based iot platform for smart cities," *Journal of Parallel and Distributed Computing*, vol. 132, pp. 250–261, 2019.

[24] X. Liang, J. Zhao, S. Shetty, J. Liu, and D. Li, "Integrating blockchain for data sharing and collaboration in mobile healthcare applications," in *Proceedings of the International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 1–5, IEEE, 2017.

[25] "GAIA-X: Technical architecture," pp. 1–56, Federal Ministry for Economic Affairs and Climate Action, 2020.

[26] "Eclipse dataspace components," in *https://projects.eclipse.org/ projects/technology.edc*. Accessed: January 2023.

[27] W. Holfelder, A. Mayer, and T. Baumgart, "Sovereign cloud technologies for scalable data spaces," *Designing Data Spaces*, p. 419, 2022.

[28] A. Sakaino, "International collaboration between data spaces and carrier networks," in *Designing Data Spaces*, pp. 471–483, Springer, 2022.

[29] C. S. Langdon and K. Schweichhart, "Data spaces: first applications in mobility and industry," *Designing Data Spaces*, p. 493, 2022.

[30] D. Bermbach, A. Chandra, C. Krintz, A. Gokhale, A. Slominski, L. Thamsen, E. Cavalcante, T. Guo, I. Brandic, and R. Wolski, "On the future of cloud engineering," in *Proceedings of the International Conference on Cloud Engineering (IC2E)*, pp. 264–275, IEEE, 2021.

[31] A. Alonso, A. Pozo, J. M. Cantera, F. De la Vega, and J. J. Hierro,´ "Industrial data space architecture implementation using FIWARE," *Sensors*, vol. 18, no. 7, pp. 1–18, 2018.

[32] "Data sovereignty as a service (DSaaS) by sovity," in *https://sovity.de*. Accessed: January 2023.

[33] V. Karagiannis, P. A. Frangoudis, S. Dustdar, and S. Schulte, "Contextaware routing in fog computing systems," *IEEE Transactions on Cloud Computing*, 2021.

[34] B. Charyyev, E. Arslan, and M. H. Gunes, "Latency comparison of cloud datacenters and edge servers," in *Proceedings of the Global Communications Conference (GLOBECOM)*, pp. 1–6, IEEE, 2020.

[35] V. Karagiannis, P. Chatzimisios, F. Vazquez-Gallego, and J. AlonsoZarate, "A survey on application layer protocols for the internet of things," *ICAS Transaction on IoT and Cloud Computing*, vol. 3, no. 1, pp. 11–17, 2015.

[36] M. Xu, Z. Fu, X. Ma, L. Zhang, Y. Li, F. Qian, S. Wang, K. Li, J. Yang, and X. Liu, "From cloud to edge: a first look at public edge platforms," in *Proceedings of the Internet Measurement Conference (IMC)*, pp. 37– 53, ACM, 2021.

[37] D. Loghin, L. Ramapantulu, and Y. M. Teo, "Towards analyzing the performance of hybrid edge-cloud processing," in *Proceedings of the International Conference on Edge Computing (EDGE)*, pp. 87–94, IEEE, 2019.

[38] V. Karagiannis and S. Schulte, "Distributed algorithms based on proximity for self-organizing fog computing systems," *Pervasive and Mobile Computing*, vol. 71, p. 101316, 2021.

[39] M. Barzegaran, N. Desai, J. Qian, and P. Pop, "Electric drives as fog nodes in a fog computing-based industrial use case," *The Journal of Engineering*, vol. 2021, no. 12, pp. 745–761, 2021.

[40] V. Karagiannis and S. Schulte, "edgerouting: Using compute nodes in proximity to route IoT data," *IEEE access*, vol. 9, pp. 105841–105858, 2021.

[41] "Gaia-x hubs," in *https://gaia-x.eu/who-we-are/hubs/*. Accessed: January 2023.

[42] R. Mathumitha, P. Rathika, and K. Manimala, "Big data analytics and visualization of residential electricity consumption behavior based on smart meter data," in *Proceedings of the International Conference on Breakthrough in Heuristics And Reciprocation of Advanced Technologies (BHARAT)*, pp. 166–171, IEEE, 2022.

[43] V. Karagiannis, "Area limitations on smart grid computer networks," *International Journal of Wireless and Microwave Technologies (IJWMT)*, vol. 6, no. 3, pp. 71–78, 2016.

[44] "Refit data," in *https://repository.lboro.ac.uk/articles/dataset/REFIT Smart Home dataset/2070091/1*. Accessed: January 2023.

[45] K. Dolui and S. K. Datta, "Comparison of edge computing implementations: Fog computing, cloudlet and mobile edge computing," in *Proceedings of the Global Internet of Things Summit (GIoTS)*, pp. 1–6, IEEE, 2017.

[46] National IT Summit. 2015. *Leitplanken Digitaler Souveränität*. Berlin. Accessed from https://www.de.digital/DIGITAL/Redaktion/DE/Downloads/it-gipfel-2015-leitplanken-digitaler-souveraenitaet.pdf?__blob=publicationFile&v=1.

[47] Digital Summit. 2018. *Digitale Souveränität und Künstliche Intelligenz – Voraussetzungen, Verantwortlichkeiten und Handlungsempfehlungen*. Berlin. Accessed from https://www.de.digital/DIGITAL/Redaktion/DE/Digital-Gipfel/Download/2018/p2-digitale-souveraenitaet-und-kuenstliche-intelligenz.pdf.

[48] Schauf, Thomas. April 2012. *Das Internet als Netzwerk von Ökosystemen: Weniger Offenheit, mehr Konzentration?* Policy Essay: Stiftung Neue Verantwortung, Berlin.

[49] Baums, Ansgar. 2015. "Analyse – Was sind digitale Plattformen." In *Industrie 4.0: Wie digitale Plattformen unsere Wirtschaft verändert – und wie die Politik gestalten kann*, edited by Ansgar Baums, Michael Schössler, and Bill Scott, 14–25. Kompendium Digitale Standortpolitik, Vol. 2. Berlin. Accessed from http://plattform-maerkte.de/wp-content/uploads/2015/11/Kompendium-High.pdf.

[50] Dobusch, Leonhard, Falk Bahr, Thorsten Dapp, Marcel Grzegorzek, Volker Kerst, Ralf Meinberg, Matthias Rehse, Jan Sänger, Thomas Schauf, and Hans Tillmann. 2012. *Schönes neues Internet? Chancen und Risiken für Innovation in digitalen Ökosystemen*. Policy Brief: Stiftung Neue Verantwortung, Berlin. Accessed from https://www.stiftung-nv.de/sites/default/files/12_04_policy_brief_the_business_web_20120824_final.pdf.