Vasileios Karagiannis, "Data Sovereignty and Compliance in the Computing Continuum", in *International Conference on Future Internet of Things and Cloud (FiCloud)*, IEEE, 2024, DOI: 10.1109/FiCloud62933.2024.00027

Data Sovereignty and Compliance in the Computing Continuum

Vasileios Karagiannis
Center for Digital Safety & Security
Austrian Institute of Technology
Vienna, Austria
vasileios.karagiannis@ait.ac.at

***Abstract*—The surge of data generated by the Internet of Things (IoT) has driven the evolution of different computing layers such as edge, fog, and cloud. More recently, the inherent need for IoT data management and processing within and among these layers has set the stage for the novel concept of the computing continuum. The computing continuum describes a system that seamlessly spans from the IoT through the edge, the fog, and the cloud, facilitating a fluid and dynamic allocation of computing resources tailored to IoT data workflows. Despite its innovative approach to resource utilization, the computing continuum may face significant challenges in aligning with emerging legislative measures governing data sharing and privacy. In fact, most related approaches in the literature do not consider the implications of new data protection laws. This can jeopardize the applicability of such approaches to real-world scenarios. To address this problem, we propose a system that integrates access and usage control policies to enable compliance with data protection regulations while transferring the data across the continuum. Furthermore, we highlight real-world use cases that can benefit from this approach, for example, related to data management in the context of energy analytics and environmental monitoring, and we demonstrate the potential impact of the proposed system.**

***Index Terms*—Computing Continuum, Internet of Things, Data Management, Data Sovereignty, Data Spaces, Climate Change**

I. Introduction

The computing continuum represents a novel paradigm in computing systems, that is designed to manage the vast amounts of data generated by users, systems, and the IoT [1]–[4]. It encompasses a spectrum of computational resources in IoT devices, edge, fog, cloud, and sky computing nodes, providing efficient processing of the data, and enabling a gamut of use cases in the context of smart cities, autonomous vehicles, and sustainable environments, among others. To this end, the computing continuum focuses on various facets of resource utilization (such as resource provisioning and allocation as well as service scheduling and execution), and data management such as data ingestion, storage, governance, and sharing. Data management is particularly critical because it focuses on ensuring that the data with appropriate permissions is available at the right place and at the right time, thereby facilitating a data lifecycle that upholds service quality, and adheres to privacy protection regulations [5]. This data-centric perspective is paramount to support the volume, velocity, and variety of the generated IoT data, and to enable automated decision-making across distributed computing environments.

The current literature on computing continuum approaches describes robust systems for data management and processing [4]. However, most of this literature does not account for the stringent privacy requirements imposed by new data privacy legislation [6]. This legislation often regulates cross-border data transfers, and mandates explicit user consent for data processing operations, aiming to protect user privacy and ensure data sovereignty. Despite this, most existing approaches frequently overlook these legal requirements, which is a significant problem in the advancement of computing continuum systems. Consequently, the applicability of such systems can be challenged due to legal compliance concerns. Also, the deployment of systems that operate at the edge (e.g., on user apps), the fog (e.g., on small data centers in the vicinity), the cloud (e.g., on large data centers), or the sky (e.g., on multiple data centers) might lead to compliance violations. In fact, many severe penalties have already been issued for potential breaches in recent years, such as the record fine of 1.2 billion EUR to Meta in 2023 for moving data cross-border, and the fine of 746 million EUR to Amazon in 2021 for not obtaining explicit user consent for data processing [7]. Therefore, as data moves fluidly across borders from IoT devices and users to edge and cloud nodes, and as systems increasingly automate decision-making, the necessity for integrating legal compliance into the continuum becomes essential.

To address this problem, we propose a computing continuum system that aligns with data sovereignty principles to handle the transfer of data across the continuum. This system incorporates access and usage control policies specifying permissions, prohibitions, and constraints regarding the terms of use of the data. This ensures that crucial information about the data (e.g., origin, processing consent, etc.) is not lost while moving the data across the computing layers of the continuum. As a result, data consumers can examine a policy before accepting data for processing, and also keep the policy as proof for audit purposes [8]. To automate this process, policies are formulated using the machine-readable Open Digital Rights Language (ODRL) [9]. Similar datasets can use the same policies, and data streams only need one policy. This way, the proposed system aims to provide native support for conveying crucial information that is necessary to ensure legal compliance while managing data within the computing continuum.

Specifically, the contributions of this work include: 1) Motivating the need to integrate data policies in the computing continuum. 2) Designing a computing continuum system with native support for data policies. 3) Identifying technological and societal challenges that may pose obstacles to successful realization. 4) Presenting preliminary results considering applicable use cases showcasing the potential impact.

The remainder of this paper is structured as follows: Section II outlines related work, and Section III presents the proposed system and associated challenges. Afterward, Section IV demonstrates applicable use cases and shows relevant results. Finally, Section V concludes this paper and suggests future work on this topic.

II. Related Work

In the literature, various approaches have been proposed for building computing systems that operate over the computing continuum, with a primary focus on resource allocation and service execution [10]. For example, Dustdar et al. [11] emphasize the need for designing novel computing systems that run concurrently in multiple computing layers, and propose a methodology to manage the underlying computing resources. Rosendo et al. [12] introduce a platform for running applications across the computing layers of the continuum, supporting end-to-end performance analysis for better utilization of the underlying infrastructure. Ferrer et al. [13] present an architecture and a resource management system for service provisioning in the continuum which may also include temporal swarms of resources. Finally, to cope with the rapid growth of data from digital sources, AbdelBaky et al. [14] propose a system that combines distributed resources and services to support novel data-driven application workflows. The authors also highlight that these workflows may need to react to dynamic availabilities and locations. While such approaches constitute crucial contributions to the literature of the computing continuum, they do not cover compliance with recent regulations, which is the focus of our work.

Additionally, there are approaches focusing on the use of data within the continuum. For instance, Roman et al. [15] explore new opportunities for managing data in the continuum, and discuss the data lifecycle and the associated challenges. Syrigos et al. [16] introduce a framework aimed at efficient data ingestion and training in machine learning applications. Ayed et al. [17] present methodologies for secure data management with a particular focus on sensitive data. Finally, Kimovski et al. [18] propose an approach for improving the management of data in the continuum by allowing domain experts to participate in the monitoring and analysis of the resources. Overall, such approaches offer significant contributions to the field of data management in the computing continuum. However, they do not address legal compliance requirements that are within the scope of the work at hand.

Approaches that consider data sovereignty and legal compliance also exist, but they focus primarily on a single computing layer, often the cloud. Kotka et al. [19] present a detailed view on transferring data to the cloud while taking into account data protection and data sovereignty concerns. Geisler et al. [20] provide an overview of data ecosystems and the associated challenges of achieving data sovereignty and privacy. Firdausy et al. [21] design a reference architecture to aid enterprises in implementing data sovereignty principles. In addition, there are related contributions from organizations that aim at enhancing the understanding of sovereign data. Notably, the Alliance of IoT and Edge Computing Innovation (AIOTI) has produced reports on integrating sovereign data with edge computing. The International Data Spaces Association (IDSA) and Gaia-X have published reports that describe a cloud-based infrastructure, also referred to as a data space, for realizing data sovereignty. Based on these reports, various works follow the data space method, often with a cloud-based connector component that transfers the data [9], [22]. While such works highlight the importance of data sovereignty, they do not focus on enabling computing continuum systems to comply with data protection laws, which is the goal of this paper.

In summary, the review of related literature reveals a growing number of approaches that focus on conceptualizing and implementing computing continuum systems. However, enabling these systems to achieve compliance with modern privacy legislation still remains a challenging endeavor. For this reason, this paper focuses on compliance aspects in the computing continuum and proposes a system that is driven by data policies. Additionally, we introduce applicable use cases and we provide preliminary findings which emphasize the potential of the proposed approach.

III. The Computing Continuum

In this section, we present the computing continuum as a data space designed to uphold data sovereignty, and to foster compliance with contemporary legislation. To this end, Section III-A identifies the fundamental requirements of the system, and Section III-B provides a system overview. Afterward, Section III-C examines the potential challenges that may need to be addressed on the path to realization.

A. System Requirements

As discussed in Section I, current computing systems might struggle to comply with two prime challenges of data protection legislation: obtaining explicit consent from users for data access and processing, and transferring data across the continuum (and over national borders). To address these challenges, we focus on aspects related to data governance and transfer. To improve these aspects in modern computing systems while considering applicable legislation, we identify requirements Req. 1–6 that need to be met:

1) *Data Sovereignty*: Refers to the principle that data is subject to the laws of the country where it is collected and the rules of the data owners who decide how the data can be used, by whom, and for what purpose. The system shall ensure that data is handled in compliance with this principle. This is crucial because unauthorized data access, usage, and transfer within and across borders can conflict with data protection laws.

2) *Trust and Openness*: The system shall foster trust among participants by providing clear information regarding data policies and handling. When sharing data, data owners shall have guarantees that their data will be used according to legal and user-defined constraints. This is important to assure participants of the proper use of their data.
3) *Semantic Interoperability*: Refers to the ability of different systems to interpret and understand the meaning of exchanged information. Given the continuum's diverse computing layers, semantic interoperability becomes crucial to ensure accurate interpretation between different system components.
4) *Audit Trail and Reporting*: Refers to tracking and documenting data access and sharing for accountability and compliance purposes. The system shall provide audit trail mechanisms that record all data transactions, with access to these records limited to the transaction participants to ensure their privacy. Such an audit trail is vital for regulatory accountability in data handling practices.
5) *Domain Interoperability*: The computing continuum may process data from various domains, both individually and in combination. Combined processing of datasets from different domains, in particular, necessitates the integration of diverse data along with their (possibly conflicting) legal requirements. Thus, ensuring interoperability of data across domains is crucial for maintaining compliance with domain-specific regulations in the continuum.
6) *Decentralization*: The computing continuum is inherently distributed and decentralized. Therefore, any proposed enhancements shall adhere to these attributes, aligning with the continuum's fundamental principles. This ensures equality, preventing any single entity from exerting undue control or gaining unauthorized access to data.

B. System Overview

Our system aligns with state-of-the-art computing continuum systems from the literature [11], [23], [24]. Nevertheless, this paper describes additional components and mechanisms that aid in meeting the goals outlined in Section III-A. A high-level view of the computing continuum is depicted in Fig. 1. As shown in this figure, the continuum encompasses a range of computing resources, extending from IoT and user devices to the edge, fog, and cloud layers. While the IoT and user devices can function both as data sources and computing resources, the edge, fog, and cloud layers are primarily utilized for computing tasks, including processing, storage, and management operations.

Typically, the IoT comprises resource-constrained wired and wireless devices equipped with sensors and actuators that form sensor networks in areas of interest, such as residential regions, agricultural fields, and coastlines. The raw data generated, which is crude but potentially valuable, undergoes processing at various stages. Nearby edge nodes facilitate low-latency processing for time-sensitive tasks requiring immediate response, e.g., energy monitoring for predictive failsafe mechanisms that avert hazards. These edge nodes may consist of on-site computers and base stations in close network proximity, typically one hop away from the data sources [25]. A few hops away, fog nodes offer more substantial computing power with a small delay. Since these nodes are also in proximity, bottlenecks are unlikely, making tasks related to analysis and decision-making well-suited. Fog nodes may include dedicated on-site computers that aggregate computations from multiple data sources, and network devices with available computing resources. Farther away from the fog, cloud nodes provide virtually unlimited resources, appropriate for mass data storage, processing, and training of large-scale machine learning models. Due to their potentially remote location and their primary role in mass processing, cloud nodes may experience bottlenecks and high latency.

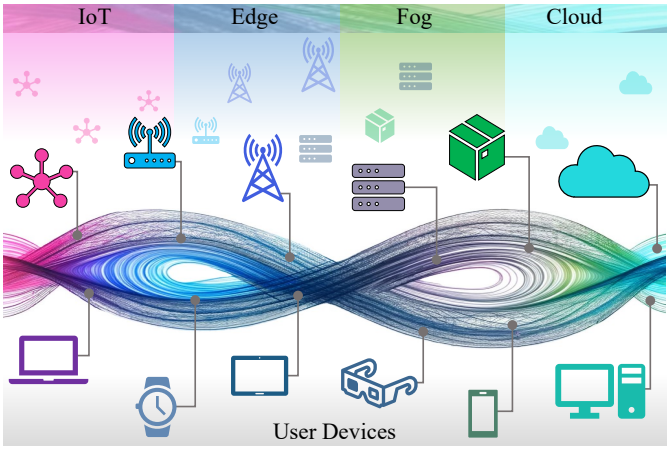

Fig. 1: The computing continuum includes various computing layers and devices interacting with each other seamlessly.

Users are also integrated into the continuum via various user devices such as wearables, virtual reality headsets, smartphones and laptops, which may connect to any of the computing layers. As a result, data can flow seamlessly between layers starting at any point in the network and heading toward any other. Accordingly, Fig. 1 illustrates the continuum as a continuous flow of data moving across the computing layers (symbolized by the infinity sign ∞). Notably, as shown in the figure, a single device type may belong to many layers (e.g., at the edge and the fog) since a layer is defined not only by device types but also by their network location and proximity to data sources. As data flows and undergoes continuous processing across nodes in different locations within the continuum, tracking data ownership, privacy, and provenance becomes exceedingly difficult.

For this reason, we propose the system of Fig. 2 which includes additional components and actors. Specifically, this system comprises the following actors:

- The *data owner* is an entity (individual or organization) that owns a dataset.
- The *data provider* is an entity that possesses a dataset acquired from a data owner (or a data provider) and distributes it.

- The *data consumer* is an entity that acquires a dataset from a data provider (or owner) for processing purposes.
- The *federator* is an entity that maintains federated services, which support the operation of the system.

Every participant in the system assumes at least one of these roles, possibly more. Each dataset exchanged within the system is accompanied by a policy. The data *policy* outlines the dataset's terms of use by specifying permissions, prohibitions, and constraints using ODRL. Created by the data owner, the policy reflects the origin of the data, country-specific regulations, and the owner's additional access/usage rules, e.g., the consent for processing. When a data provider redistributes a dataset, the same policy of the data owner applies.

Data can flow freely across the continuum provided it remains under the control of a single participant and within a country's boundaries. If the situation differs, the dataset's policy must be agreed upon between the provider and the consumer through a transaction, before exchanging the data. A data exchange can be initiated by any participant. For example, when the consumer requests a dataset, the provider responds with the dataset's policy digitally signed. The consumer then reviews the policy and, if in agreement, signs it digitally before returning it to the provider. Following this, the provider dispatches the dataset to the consumer. Through this process, both participants obtain evidence of the policy agreement (i.e., the signed policy), while the public keys (of the signatures) can be used for authentication. As data is propagated across the continuum, it is always accompanied by the originating owner‘s policy. Should the policy restrict processing, e.g., because a location requirement is not met, the consumer can recognize this restriction and opt not to sign, thereby not receiving the data. This process occurs at least once for every type of data. For streaming data, for example, which is common in the IoT, the process takes place once and the same policy applies to the whole stream. This process aims to ensure that the policy is communicated, and aids in meeting Req. 1 and 2 of Section III-A.

To enhance trust among the participants, all transactions are validated and recorded using the *Transaction Logger* which is a federated service operated by federators. For a transaction to be logged, its initiator must be a verified entity. Verification is conducted through a credible self-sovereign identity or via a custom identity provider (implemented by the logger) that issues valid verifiable credentials following an onboarding process. Participants have the option to forgo transaction logging, to log their transactions voluntarily, or to require that their data exchange partners also engage in logging. While not mandatory, transaction logging constitutes an additional layer of protection and serves as evidence for accountability and conflict resolution purposes. This also contributes to the fulfillment of Req. 2 and 4 in Section III-A.

As multiple participants exchange and process data from various domains within the continuum, a discovery mechanism for data becomes increasingly useful. The *Metadata Registry* is a federated service that enables data providers to submit metadata describing their datasets (optionally). Consumers can search this registry to identify datasets of interest and request them directly from the providers. Given that datasets can be large and complex, the relevant data models and semantics may be essential for efficient processing. Thus, the *Semantic Registry*, another federated service, allows providers to submit the metamodels and semantic details of their datasets. Consumers can use the Metadata Registry to find the identifier and host address of the desired datasets. Subsequently, with the identifier, consumers can retrieve the corresponding data model from the Semantic Registry. Afterward, a consumer can use the host address to request a dataset (specified by the identifier) from the provider. To facilitate interoperability across different domains, the Semantic Registry also accepts mappings. The mappings provide correspondences and conversion functions between different data models. Hence, both registries play an important role in meeting Req. 3 and 5 as specified in Section III-A. Importantly, the federated services handle only metadata and control data, whereas the actual datasets are exchanged exclusively between participants. Furthermore, these services are collaboratively operated by federators in a distributed fashion, avoiding any centralized components. This is vital for meeting Req. 6.

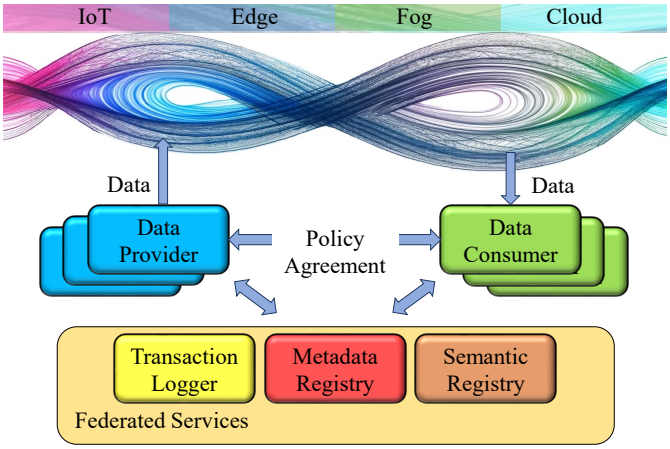

Fig. 2: Data sharing within the computing continuum involves establishing policies and utilizing federated services.

Overall, devices and computing nodes within the continuum may act as providers and/or consumers as depicted in Fig. 2, exchanging data automatically. This exchange, however, adheres to a defined process that ensures that both providers and consumers are safeguarded by the data policies and corresponding logs. Since the Transaction Logger verifies participants, access to the logs is restricted to those involved in a transaction in order to protect the participants' privacy. Notably, when data is consumed and a new derivative dataset is created, this new derivative dataset inherits the policy of the original data. The consumer, having created the derivative dataset, defines a new policy appending the inherited policy which still applies. Consequently, consumers have the responsibility to carefully select the origin datasets, because the origin data policies affect the policy of the derivatives.

C. Open Challenges

Since compliance with regulations is a societal concept, the challenges associated with the proposed approach extend beyond technological to societal aspects. Some of the key open challenges are discussed below:

Software Implementation: The proposed approach may require a significant engineering effort to be implemented. This includes the logic of the components, the processes, the interfaces and the data models. Interestingly, existing open-source implementations can contribute to this effort. For example, the Connector of the Eclipse Foundation aims at exchanging data policies using ODRL, and the Tractus-X project provides registries for metadata and semantics for the automotive industry [26]. Thus, while additional effort is certainly needed, some building blocks do exist.

Usage control: Enforcing adherence to agreed-upon data policies can be challenging, especially when it comes to detecting non-compliance of data consumers. In some cases, policy compliance can be enforced. For example, a user-defined policy of allowing access to real-time data from a database for one month can be implemented with access credentials that expire after 30 days. However, if a policy prohibits specific processing methods, e.g., machine learning-based predictive methods, ensuring compliance becomes more difficult. One solution might be for the provider to process the consumer's computations and return to the consumer only the result. This does not allow the consumer to access the data, thus avoiding further data processing and potential policy violations. However, this method has its drawbacks, as the standalone result may have limited value for the consumer, and lack credibility. Overcoming such challenges may require the involvement of policymakers who can hold participants accountable for their actions. To allow this, the system has to maintain records of transactions and agree-upon policies, as proposed by our approach.

Conflicting policies: Formulating policies can be straightforward in some cases, but it can become complex when reflecting legislation. There may also be instances that a dataset, being derivative of consecutive processing by multiple participants, is subject to several policies from different countries, potentially with conflicting terms. In general, when in conflict, the most restrictive policy should prevail, although conflicts might become more complicated to resolve. The issue of conflicting policies is encountered in many domains and typically requires the collaboration of policymakers to resolve. Notably, significant progress in formulating and applying policies to data has been made in the context of digital content access, with ODRL being a very promising approach for defining clear data policies.

Data ownership: The ownership of derivative data generated by a consumer using a provider's dataset as a base can be controversial. Specifically, it can be debated whether the ownership should belong solely to the consumer or include the provider as well. This topic is similarly controversial in the context of copyright laws. In this work, we posit that the focus should be on inheriting the policies of all the base datasets to ensure compliance with the legal requirements. The nominal ownership might not have a significant impact on the system's operation (although in specific cases, this might change).

Legal accountability: The goal of our approach is to convey policies and help data consumers avoid violating regulations. However, the responsibility of formulating these policies lies with the data owners/providers. In case the policies are wrong, it is difficult to hold the data providers accountable, when it is the consumers that unlawfully process the data. In addition, while the system provides logs about the policy agreements, whether these logs have legal standing is a societal matter. Thus, it is important to align the computing continuum with policymakers, so that the proposed processes have meaning in technological—and legal contexts.

Incentives: The proposed approach, focusing on data sharing, might require additional mechanisms to keep all the actors motivated to operate and utilize the system components. However, as data protection legislation becomes increasingly stringent, the demand for a computing continuum equipped with integrated data policies and audit trails becomes more essential. Further motivation for sustained participation can be fostered by introducing incentive mechanisms, such as those derived from game theory, pricing models, and marketplaces.

IV. Use Cases

To demonstrate the applicability of our approach in real-world scenarios, in this section, we present use cases that stand to benefit from the proposed system. To this end, Section IV-A examines the exchange of energy consumption data from residential areas in the context of performing energy analytics, and Section IV-B focuses on the collection and processing of environmental monitoring data.

A. Energy Consumption Data

Residential areas are significant contributors to global energy use, exerting a substantial impact on the world's carbon footprint. Collecting and analyzing energy consumption data from these areas can lead to the identification of key regions for improvement and opportunities for renewable energy integration and crucial energy analytics. This data provides insights into daily energy usage peaks, seasonal fluctuations, and the effectiveness of energy-saving measures [27]. Analyzing these patterns using various techniques [28], [29], can contribute to the reduction of greenhouse gas emissions, thereby aiding in the mitigation of climate change effects. Therefore, the collection and analysis of residential energy consumption data is also crucial for climate change research.

Despite the importance of this data, household energy consumption information is typically not readily available. Until recently, access to energy consumption data was exclusive to the energy providers (for most regions) who used this data primarily for billing. However, with the emergence of data as an enabler of optimizations, new legislation mandates that energy providers grant access to energy consumption data to both their customers and third-party entities [27]. Examples

```
1. {
2.   "@context":"https://www.w3.org/ns/odrl.jsonld",
3.   "id": "http://policies.com/policy:1234",
4.   "type": "Set",
5.   "permission": [{
6.     "target": "http://registry.com/resource/123",
7.     "action": "read",                \\read
8.     "constraint": [{
9.       "leftOperand": "dateTime",     \\for date
10.      "operator": "lt",              \\less than
11.      "rightOperand": "2026-01-01"   \\year 2026
12.    }]
13.  }]
14. }
```

Fig. 3: Example of an ODRL policy in JSON-LD format.

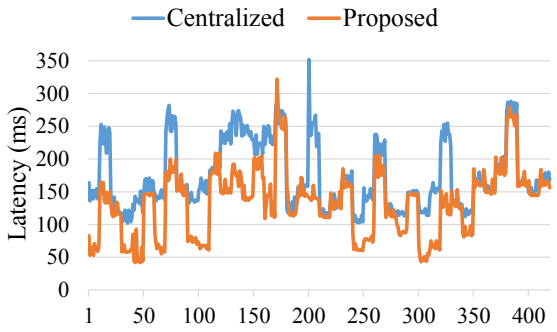

Fig. 4: Latency of data sharing with policy agreements.

of such legislation include the European Directive 2019/944, California's Decision (D.) 14-05-016, and Michigan's Administrative Rules R 460.101 to R 460.169, all regulating access to energy data. To protect customer privacy, third-party entities may access this data only with the explicit consent of the customer, as stipulated by the EU Directive, for instance.

To comply with such laws, energy providers have developed digital platforms that allow third-party access to customer data with the requisite customer consent. Examples of such platforms are, for instance, EDA in Austria, Enedis in France, and Datadis in Spain. Nevertheless, these platforms lack support for flexible policies to represent legislative measures. As a result, if the customer's consent is not effectively integrated with the data when a dataset is going through the various computing layers, the legality of any subsequent data processing within the computing continuum could be questioned. This problem hinders the processing of energy data within the continuum and obstructs its potential for energy analytics.

The proposed approach can tackle this problem using the presented components and processes. The energy provider, having the energy data, formulates an initial policy that incorporates both its own requirements and the local legal regulations. Through the provider's platform, the customer can incorporate additional terms of use (if any). Subsequently, third parties can request data access from the provider that allows access only after a third party agrees with the policy. All transactions are logged, enabling each participant to verify their adherence to the agreed-upon policy at any moment.

In the context of this work, we implement a preliminary prototype (in Java) of a data provider and a data consumer that agree to a policy before exchanging data. The provider's policy, formulated in ODRL using JSON-LD, allows the use of the data until the end of the year 2025, as shown in Fig. 3. We deploy instances of providers and consumers on computing nodes across Europe (in Spain, France, England, Belgium, Poland, Italy, and the Netherlands) using Google Cloud Services, and we configure these participants to exchange energy data from a publicly available energy dataset [9]. Notably, based on the proposed approach, the participants exchange data directly without centralized coordination. The experiments indicate that the policy agreement process adds negligible computational overhead and only a few milliseconds (ms) of delay. Furthermore, we implement an alternative approach using a centralized cloud node (in Germany), which performs the policy agreements and data transfers between all the distributed providers and consumers, as discussed in related work [22], [30], [31]. This approach serves as a baseline for preliminary latency experiments and comparisons.

The latency of data transfers for both approaches is depicted in Fig. 4. For these results, we conduct 420 experiments. In each experiment, a randomly chosen participant acts as the consumer and requests data from a randomly selected provider twice: once using the proposed approach (i.e., direct consumer-provider communication), and once using the baseline (i.e., consumer-provider communication via the centralized cloud node). Based on the results, a clear pattern of similar or lower latency is observed. This indicates that the proposed approach performs better when providers and consumers exchange data in a distributed manner which advocates the suitability of our approach for the computing continuum that is inherently a highly distributed paradigm [11]. Notably, these results are the outcome of using a dataset with quarter-hourly energy measurements of one month (with a size of 225 kilobytes).

Such a data-sharing system with integrated policies can have significant implications for the energy analytics market. Currently, most customers have access only to energy analytics services offered by their corresponding energy providers. Such services are typically very limited in functionality, e.g., basic visualizations of energy consumption values [27]. Allowing third-party companies to access energy data at scale, and enter the market of energy data services can create conditions of perfect competition. In this case, customers are no longer limited to the energy services of their energy providers, but instead, they can share their data and use services by any service provider. Notably, perfect competition leads to innovation and reduced operational costs. Subsequently, this can facilitate the development of innovative services that increase awareness about energy consumption. Such services may include consumption recommendation systems, off-peak pricing, and in-house/grid optimizations, among others.

B. Environmental Monitoring Data

Data from environmental monitoring can be considered of significant importance as it provides key indicators regarding

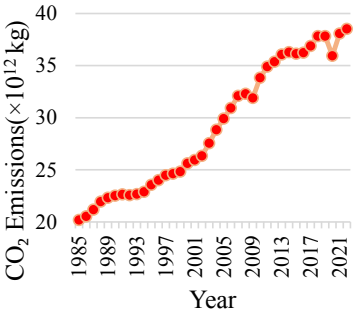

Fig. 5: Global CO_2 emissions (1985–2022).

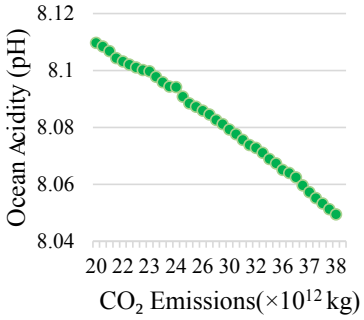

Fig. 6: Ocean acidity by CO_2 emissions (1985–2022).

the health of our planet. For example, global emission datasets track the concentration of CO_2 which is a major contributor to global warming. Additionally, ocean acidity data can indicate the increasing absorption of CO_2 by the oceans, leading to acidification. Understanding the trends in such datasets is vital for forecasting future climate change scenarios and guiding the implementation of appropriate countermeasures via policy decisions.

While crucial for climate change research, environmental monitoring data is often subject to various data protection laws that may differ according to the origin of the data. For instance, marine data collected within the European Union may fall under the General Data Protection Regulation, the Infrastructure for Spatial Information, the Marine Strategy Framework Directive, and other nation-specific regulations [32]. Therefore, unrestricted processing of this data within the computing continuum, and generation of derivative datasets while disregarding these regulations can result in violations. To prevent this, datasets governed by extensive legislation can be accompanied by policies, as proposed in our approach. When such datasets cross national borders or change organization, all parties involved need to ensure proper data handling. The proposed system facilitates this by enabling data transfers that incorporate integrated policies and audit trails.

To further highlight the significance of environmental monitoring data in climate change research, we examine closely two datasets with measurements from the past 30 years: the global CO_2 emissions, and the average ocean acidity. Notably, the availability of such datasets often results from special agreements between agencies. For instance, the CO_2 emissions dataset is made available through an agreement between the Joint Research Centre and the International Energy Agency [33], while the ocean acidity is disseminated by Europe's Copernicus Marine Service [34].

Enabling the seamless processing of such datasets within the computing continuum, as our approach intends, can result in a high degree of automation for collecting, processing, and interpreting environmental data. To demonstrate the potential of automating such processes, we analyze the two environmental datasets. Fig. 5 shows the global CO_2 emissions per year. Fig. 6 correlates each emission value with the ocean acidity of the same year. Interestingly, Fig. 6 shows that when the emissions increase, the acidity decreases, which is the result of the absorption of the CO_2 by the ocean, also known as ocean acidification. Notably, Fig. 5 has an out-of-trend low value in the year 2021, which is very likely due to the COVID-19 pandemic and the reduced manufacturing production.

Fig. 6 shows data that exhibits a clear linear pattern. Because of this, we perform a linear regression analysis with the acidity as the independent variable and the emissions as the dependent variable to determine the strength of their relationship. The analysis produces a coefficient of determination $R^2 = 0.94$ indicating a strong linear relationship, significance F ≈ 0 suggesting statistical significance, and standard error of 0.004 showing that the regression line closely fits the data. The regression equation is $acidity = -0.003 * emissions + 8$ (I). Notably, ocean life can be sensitive to changes in the acidity. For instance, some marine fish species, like the orange clownfish, can struggle to survive acidity levels of 7.8 pH or lower due to sensory impairment [35], [36]. Based on I, this pH level corresponds to emissions of approximately 100×10^{12} kilograms (kg) (II).

To estimate the year when this amount of emissions might occur, we perform a regression analysis with the emissions as the independent variable and the year as the dependent variable (data from Fig. 5). The results show a coefficient of determination $R^2 = 0.96$ indicating a strong linear relationship, significance F ≈ 0 suggesting statistical significance, and standard error of 1.18 showing a close fit of the regression line to the data. The regression function is $emissions = 0.55 * year - 1070$ (III). According to this function, 100×10^{12} kg of CO_2 emissions from (II) will be reached by the year 2127. Thus, that year the orange clownfish might start experiencing the side effects of climate change due to sensory impairment.

Based on the proposed system, such processes can be fully automated within the computing continuum. Integrated sensors for air quality, satellite imagery, and measurements from vessels can be configured to flow seamlessly through the computing layers governed by integrated policies, to establish monitoring systems that track a variety of environmental parameters. Furthermore, these systems can be linked to the

automatic generation of reports assessing the condition of critical habitats and evaluating the impacts of countermeasures, which can significantly advance the efforts towards climate change mitigation.

V. Conclusion

This paper conceptualizes the computing continuum as a system that natively supports policies allowing data consumers to adhere to data protection laws. This is motivated by the need to enhance the practical applicability of computing continuum systems in real-world scenarios. Moreover, we identify related open challenges that may need to be addressed, and we demonstrate the potential of our system by showcasing applicable use cases. Specifically, we highlight energy analytics and environmental monitoring as use cases that stand to benefit from our approach. Since compliance and sovereignty become increasingly essential in the continuum, in the future, we plan to focus on overcoming the identified challenges in order to further advance computing continuum systems.

References

[1] M. Bajer, "Iot for smart buildings-long awaited revolution or lean evolution," in *International Conference on Future Internet of Things and Cloud (FiCloud)*, pp. 149–154, IEEE, 2018.

[2] V. Karagiannis, P. Chatzimisios, F. Vazquez-Gallego, and J. Alonso-Zarate, "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.

[3] M. Younas, I. Awan, G. Ghinea, and T.-M. Grønli, "New developments in cloud and iot," 2018.

[4] P. K. Donta and S. Dustdar, "The promising role of representation learning for distributed computing continuum systems," in *International Conference on Service-Oriented System Engineering (SOSE)*, pp. 126–132, IEEE, 2022.

[5] F. Li, B. C. Ooi, M. T. Özsu, and S. Wu, "Distributed data management using mapreduce," *ACM Computing Surveys*, vol. 46, no. 3, pp. 1–42, 2014.

[6] A. Aljeraisy, M. Barati, O. Rana, and C. Perera, "Privacy laws and privacy by design schemes for the internet of things: A developer's perspective," *ACM Computing Surveys*, vol. 54, no. 5, pp. 1–38, 2021.

[7] "Data Protection Fines," in *https://tinyurl.com/37tnsume and https://tinyurl.com/24e964ty*, 2024. Accessed: 3/2024.

[8] J. Xu, N. Hong, Z. Xu, Z. Zhao, C. Wu, K. Kuang, J. Wang, M. Zhu, J. Zhou, K. Ren, X. Yang, C. Lu, J. Pei, and H. Shum, "Data-driven learning for data rights, data pricing, and privacy computing," *Engineering*, vol. 25, pp. 66–76, 2023.

[9] V. Karagiannis, A. Al-Akrawi, and O. Hödl, "Data sovereignty at the edge of the network," in *International Conference on Fog and Edge Computing (ICFEC)*, pp. 1–7, IEEE, 2023.

[10] L. Baresi, D. F. Mendonça, M. Garriga, S. Guinea, and G. Quattrocchi, "A unified model for the mobile-edge-cloud continuum," *ACM Transactions on Internet Technology (TOIT)*, vol. 19, no. 2, pp. 1–21, 2019.

[11] S. Dustdar, V. C. Pujol, and P. K. Donta, "On distributed computing continuum systems," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 4, pp. 4092–4105, 2022.

[12] D. Rosendo, P. Silva, M. Simonin, A. Costan, and G. Antoniu, "E2clab: Exploring the computing continuum through repeatable, replicable and reproducible edge-to-cloud experiments," in *International Conference on Cluster Computing (CLUSTER)*, pp. 176–186, IEEE, 2020.

[13] A. J. Ferrer, S. Becker, F. Schmidt, L. Thamsen, and O. Kao, "Towards a cognitive compute continuum: An architecture for ad-hoc self-managed swarms," in *International Symposium on Cluster, Cloud and Internet Computing (CCGrid)*, pp. 634–641, IEEE, 2021.

[14] M. AbdelBaky, M. Zou, A. R. Zamani, E. Renart, J. Diaz-Montes, and M. Parashar, "Computing in the continuum: Combining pervasive devices and services to support data-driven applications," in *International Conference on Distributed Computing Systems (ICDCS)*, pp. 1815–1824, IEEE, 2017.

[15] D. Roman, R. Prodan, N. Nikolov, A. Soylu, M. Matskin, A. Marrella, D. Kimovski, B. Elvesæter, A. Simonet-Boulogne, G. Ledakis, H. Song, F. Leotta, and E. Kharlamov, "Big data pipelines on the computing continuum: tapping the dark data," *IEEE Computer*, vol. 55, no. 11, pp. 74–84, 2022.

[16] I. Syrigos, N. Angelopoulos, and T. Korakis, "Optimization of execution for machine learning applications in the computing continuum," in *Conference on Standards for Communications and Networking (CSCN)*, pp. 118–123, IEEE, 2022.

[17] D. Ayed, E. Jaho, C. Lachner, Z. Á. Mann, R. Seidl, and M. Surridge, "Fogprotect: Protecting sensitive data in the computing continuum," in *European Conference on Service-Oriented and Cloud Computing (ESOCC)*, pp. 179–184, Springer, 2021.

[18] D. Kimovski, C. Bauer, N. Mehran, and R. Prodan, "Big data pipeline scheduling and adaptation on the computing continuum," in *Annual Computers, Software, and Applications Conference (COMPSAC)*, pp. 1153–1158, IEEE, 2022.

[19] 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 *International Conference on Theory and Practice of Electronic Governance (ICEGOV)*, pp. 103–108, ACM, 2016.

[20] S. Geisler, M.-E. Vidal, C. Cappiello, B. F. Lóscio, A. Gal, M. Jarke, M. Lenzerini, P. Missier, B. Otto, E. Paja, B. Pernici, and J. Rehof, "Knowledge-driven data ecosystems toward data transparency," *ACM Journal of Data and Information Quality (JDIQ)*, vol. 14, no. 1, 2021.

[21] 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 *Conference on Business Informatics (CBI)*, vol. 1, pp. 117–125, IEEE, 2022.

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

[23] V. Karagiannis, P. A. Frangoudis, S. Dustdar, and S. Schulte, "Context-aware routing in fog computing systems," *IEEE Transactions on Cloud Computing*, 2023.

[24] A.-N. Mays, V. Karagiannis, T. De Block, and B. Volckaert, "Federated scheduling of fog-native applications over multi-domain edge-to-cloud ecosystem," in *International Conference on Network and Service Management (CNSM)*, pp. 1–7, IEEE, 2023.

[25] V. Karagiannis and A. Papageorgiou, "Network-integrated edge computing orchestrator for application placement," in *International Conference on Network and Service Management (CNSM)*, pp. 1–5, IEEE, 2017.

[26] "Eclipse connector," in *https://tinyurl.com/2ucarvev*, 2024. Accessed: 3/2024.

[27] V. Karagiannis, S. Kashyap, N. Zechner, O. Hödl, G. Hartner, M. Llorca, T. Jamasb, S. Grünberger, M. Kurz, C. Schaffer, and S. Schulte, "A framework for enabling cloud services to leverage energy data," in *International Conference on Cloud Engineering (IC2E)*, pp. 43–50, IEEE, 2023.

[28] V. Dadvar, L. Golab, and D. Srivastava, "Exploring data using patterns: A survey," *Information Systems*, vol. 108, p. 101985, 2022.

[29] 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.

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

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

[32] V. Karagiannis, M. AL-Naday, and T. De Block, "The blue dataverse: A system for marine data sovereignty," in *World Forum on the Internet of Things (WF-IoT)*, pp. 1–6, IEEE, 2023.

[33] "Global co2 emissions dataset," in *https://doi.org/10.2760/953322*, 2024. Accessed: 3/2024.

[34] "Ocean acidity dataset," in *https://doi.org/10.48670/moi-00224*, 2024. Accessed: 3/2024.

[35] P. L. Munday, D. L. Dixson, J. M. Donelson, G. P. Jones, M. S. Pratchett, G. V. Devitsina, and K. B. Døving, "Ocean acidification impairs olfactory discrimination and homing ability of a marine fish," *Proceedings of the National Academy of Sciences*, vol. 106, no. 6, pp. 1848–1852, 2009.

[36] P. L. Munday, G. P. Jones, M. S. Pratchett, and A. J. Williams, "Climate change and the future for coral reef fishes," *Fish and Fisheries*, vol. 9, no. 3, pp. 261–285, 2008.