Data-Driven IoT Ecosystem for Cross Business Growth: An Inspiration Future Internet Model with Dataspace at the Edge

Parwinder Singh*, Nidhi†, Michail J. Beliatis‡, Mirko Presser§
Department of Business Development and Technology, Aarhus University, Herning, Denmark
Emails: *parwinder@btech.au.dk, †nidhi@btech.au.dk, ‡mibel@btech.au.dk, §mirko.presser@btech.au.dk.

Abstract*—Data is the bloodline for a business to grow, compete, and sustain in the market. It empowers businesses to build diverse services comprising innovative business models. For this, businesses must adopt an open collaboration approach, making their data and associated services available for sharing and reuse purposes, leading towards a positive and collaborative win-win business model instead of competing with each other. This creates the need for a digital ecosystem that allows data and services to be shared, reused and exchanged in a governed and secure manner. Dataspace (DS) caters to the same objective that facilitates many data operations for stakeholders, such as search, query, aggregation, federation, integration, analysis, etc., over geo-spatially distributed and diverse resources. Therefore, we propose a novel edge-enabled context-aware Dataspace model, presented for the first time in literature, as a potential solution to integrate cross-domain and cross-organization data and associated services in local or regional contexts. This model aligns with the architectural vision of the future internet model, which can create collaborative innovation and shape the futuristic industry 5.0 and beyond ecosystems. In this context, each participating organization will act as an edge that supports DS computing resource requirements and offers edge-oriented advantages in saving latency, bandwidth, and data operations near or at the data source. The model has also been validated over a local IoT edge-cluster emulated Dataspace testbed and found to fulfill the functional aspects of the proposed model.

Keywords*— *Cross Domain; Architecture; Context Aware; Data Lake; Data Space; Dataspace; IoT; Edge; Platform; Semantics.

I. Introduction

Data, in the Internet of Things (IoT) ecosystem, is an asset to active (primary) and passive (secondary) users, i.e., generated data for specific purposes can be useful for other applications based on data sharing and exploitation rights in its raw or processed form. Dataspace (DS) has emerged as a paradigm to facilitate seamless data integration from various heterogeneous data sources, including corporate databases, files, web services, IoT-oriented devices, platforms, gateways, services, etc. It administers a virtual space to pool data from various sources under its owner rights until requested access from another application or service [1].

DS expedites cross-domain data management operations and creates a unified data catalog, acting as a regulated data marketplace adhering to relevant policies for fair data usage [2]. It enables a user-friendly semantic representation of data context with built-in security and privacy measures, leading to numerous opportunities and innovative business models for different stakeholders engaged in the data life cycle and connected over the Dataspace value chain network [3]. For example, DS can enable the Pay-as-You-Go business model [4] and generate revenue from available data through its pooling, sharing, reusability, and access capabilities [5]. Figure 1 illustrates

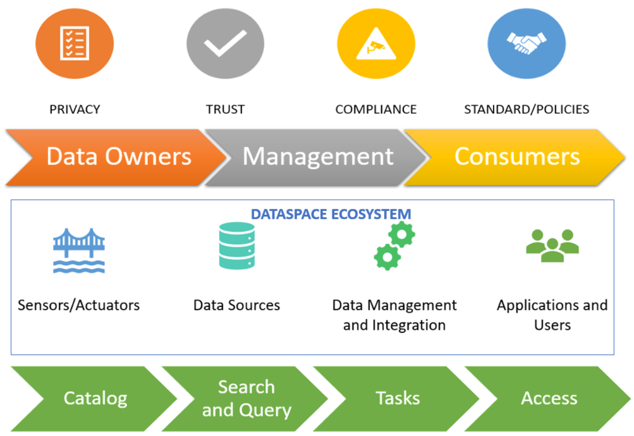

Fig. 1. Dataspace Ecosystem and Associated Players.

the interaction between different stakeholders in the DS ecosystem. However, data integration for DS faces many challenges in developing cross-domain data and service value chains. Therefore, this raises an important Research Question (RQ):

How to build DS in the local context for developing data-driven cross-domain service value-chain enablement?

The surge in connected IoT devices demands a resilient and robust future internet infrastructure to facilitate efficient data management and associated operations [6]. Therefore the role of distributed edge computing becomes more significant in supporting edge-enabled DS ecosystem [7] to address and optimize the arising challenges of security, privacy, standardized integration practices, and transforming the digital landscape towards sustainability [8]. The concept of DS revolutionizes the way we perceive and utilize data across the entire value chain, facilitating diverse services enablement and monetization opportunities that drive growth and create lasting impact. Aligning with the RQ, we have broadened the understanding of the DS concept with a focus on *how such an ecosystem can be realized at the edge* or on-premises environment, contrary to a centralized cloud facility to avail optimized latency, bandwidth, and data operations.

DS at the edge can allow data and associated services to be shared, reused, exchanged, and integrated across domains in local or regional contexts. However, realizing such a cross-domain integration ecosystem is often bundled with challenges like linked computing resources and data pool, heterogeneity, dynamic deployment context, interoperability, trust, governance, participatory motivation, etc. [3]. Therefore, it becomes critical to enable the semantic capabilities of the data to build a context-aware edge-enabled DS model. Data

 ISBN: 978-1-68558-133-6

context awareness enhances understanding, aiding discovery, quality assurance, and integration. It establishes a semantic layer for linked data within the DS ecosystem, ensuring higher data quality and reliability [7], [9]. The context-aware linked DS can enable semantic integration and harmonize the relationships within data, unlocking new insights and possibilities [3], [7]. Additionally, it will bring synergy with constantly evolving user requirements by facilitating data and technology convergence [10]. In the context of cross-domain edge (representing organization, domain, system, or service) integration empowers DS with required computing resources, availability, and convergence of technologies that enable diverse stakeholders to build a unified ecosystem for innovative business models and dynamic data-driven applications [3], [7], [11].

Therefore, this study has contributed to the semantics enablement and smart governance of the data management and associated services in future internet hyper-connected applications, particularly considering 6G and beyond [12] network ecosystems. This is achieved by identifying relevant stakeholders' common requirements, proposing and designing a Dataspace model with context-aware data processing, smart governance, and semantic adaption capabilities. In addition, a novel service artifact methodology, consisting of a service catalog and relevant toolchain, is also introduced to realize such a DS model efficiently over a distributed edge network.

The rest of the paper is given as follows: Section II will summarise relevant literature on DS and highlight key takeaways, and Section III will explain the overall methodology of this study. Further, Section IV will provide the system model, and deployment architecture framework to realize the proposed DS platform. Finally, Section V will conclude the paper.

II. Literature Review Outcomes

This section summarises the relevant literature on DS and related enabling techniques and technologies. The DS ecosystem offers a promising solution by breaking down data silos and promoting cross-domain data sharing with contextual semantics [1]. Initiatives like the International Data Space (IDS) and GAIA-X in the EU have outlined architectural frameworks and guidelines to strengthen the data economy by developing DS ecosystems [13] to facilitate seamless data integration in a larger context. StreamPipes Connect, a distributed edge-driven semantic adaptation toolbox, allows harmonizing data in Industrial IoT analytics by enabling data ingestion, sharing, and data model automation [14]. In realizing DS, addressing heterogeneity [15] is critical and can be resolved by leveraging semantics wherein ontologies represent machine-readable conceptualization of knowledge understanding at the domain level, and metadata represents a data structure at the business and technical level [5]. Thus, it is evident that metadata and ontology are essential for developing semantic information by mapping the business-level domain information to relevant technical-level information, consisting of data encapsulated entities, objects, and their inter-relationships that represent associated operations.

To build a DS ecosystem, multiple participants or entities are required. Here each entity consists of data sources and associated services with a specific or cross-domain that are geo-spatially distributed [3] and supports diverse data types or formats to represent the relevant domain-level information [3], [5]. DS essentially provides data co-existence, sharing, and reusability while promoting pay-as-you-go methods or services over the integrated data [5]. DS, in general, does not control or own the data sources, thereby, the data maintenance and administration falls under the individuals or their relevant organizational management systems [16]. Therefore, the European GAIA-X project [17] has focused on a cross-ecosystem data exchange with data sovereignty based on linking data principles. It facilitates the "common data space" concept for implementing a future "space data economy" in a cooperative business space through a common GAIA-X standard [18] supporting interoperability, portability, and data sovereignty as guided in the European data strategy [19]. Semantic modeling development tools such as Plasma are really helpful for non-technical users in providing a visual editing interface to build semantic models for DS operations [20]. These tools allow the creation, extension, and export of the semantic models and related ontologies along with relevant maintenance of knowledge graphs to annotate the datasets with semantic descriptions and convert them into unified and Resource Description Framework (RDF) standard format [21]. In IoT landscape, an edge-driven DS incorporating 'virtual sensors' allows for abstracting and mapping high-level user-driven application behaviors [22]. The user actions (in the form of HTTP verbs PUT/GET/POST, etc.) are to be reflected at the edge device, which is linked to the virtual sensor, through the application and leveraging Next Generation Service Interface - Linked Data (NGSI-LD) semantic standards information model [23].

There are also some DS-related architectural studies found in the literature. For example, [24] presents a DS testbed for maritime domain-driven data management operations which is based on a Service-Oriented Architecture (SOA) and layer-based structure emphasizing data protection and sovereignty to cater to diverse needs and support activities among multiple stakeholders. This model, however, does not address heterogeneity among various data sources. Similarly, [7] presents a Dataspace integration enablement framework based on the convergence of technologies and extending the (Cloud-Edge-Device) CED model with semantics capabilities that offer dynamic data, processing, and service context. This study has been used as the basis to define our current proposed model with a focus on context-aware DS development at the application and data management level.

Subject to limited literature about building edge-enabled DS platforms in the local context, this study contributes at the design level by proposing a distributed edge-enabled DS model with context-aware linked data and semantic adaptation capabilities.

III. Methodology

To address the RQ, we have identified the requirements based on DS stakeholders analysis [25], established methods for utilizing shared services [26], data reusability, embedding semantics in data, and creating values through context-aware linked data [27] within our local context at the Department of Business Development and Technology, Aarhus University. Our stakeholders include students, teachers, researchers, and industrial partners, where we find that *data and related operations* are the common entities among different projects. Therefore, we set a vision to extract useful information from the data semantically and collaboratively while the actors still have sovereign control over data with a readily available toolchain to perform certain semantic operations over the fusion of data in a context-aware and cross-domain manner. In this context, Figure 2 shows the value chain interaction (color-coded lines) among different stakeholders for cross-project (representing cross-domain) data-driven events and operations. This emerges as a requirement to develop a DS ecosystem in the local context to cater to diverse data management requirements. The functionality for identified requirements has been fulfilled by building a context-aware DS solution (i.e. testbed)

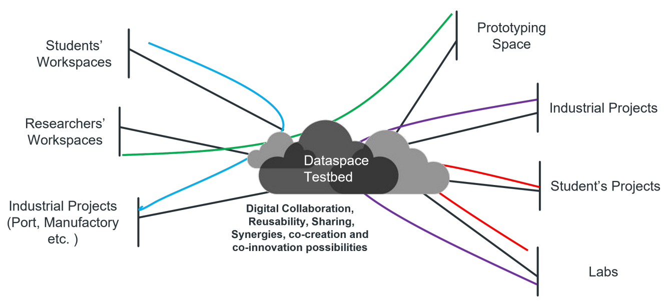

Fig. 2. Local Context Dataspace - Stakeholders and Value Chains.

following the proposed system model based on Onion architectural [28] methodology, deployment architecture [29], and selected use of toolchain as per target use case defined by the A-La-Carte (ALC) approach [30]. The solution is further validated for functional compliance against a cross-domain wind turbine supply-chain use case.

The main objective of this local context-driven DS platform is to empower hyper-connected applications and use cases in future internet-based distributed edge computing models where multiple stakeholders (dealing with different use cases, e.g., cross-lab collaboration activities, prototyping and training initiatives, external industrial projects, student education, etc.) and their data interactions will develop relevant value chains in their contextual space, as shown in Figure 2. Therefore, we proposed a semantics-driven DS model with context-aware data lake functional capabilities and realized it in our local lab environment. The next section covers the relevant details.

IV. System Model and Framework

This section proposes a reference semantic DS model implemented with context-aware and semantic adaptation capabilities to ensure that the context associated with the data under diverse DS operations enables data value in a given context and empowers data usefulness.

A. Requirements Analysis

Figure 3 illustrates high-level requirements to realize the DS ecosystem based on our stakeholder discussions, which are explained as follows:

- *Multistakeholder and Cross-Collaboration* - This indicates that the DS should support multi-tenancy operations across domains to promote collaboration while securing ownership, isolation and segregation aspects. This will enable the development of cross-domain service value chains over the data integrated in DS.
- *Monetization* - One of the main objectives for DS development is to generate monetary values from the DS integrated ecosystem by building innovative business models based on each other's data strengths. This can be the basis for a data marketplace where data and associated services generate real value and motivation.
- *Data Operations* - The system should allow data management i.e., CRUD (creation, updation, deletion, and read), operations along with federation, analytical, and visualization contextually.

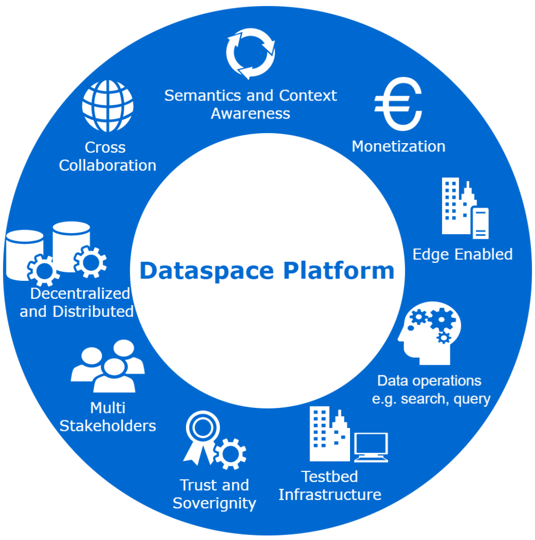

Fig. 3. High-Level Requirements for Edge Enabled Dataspace.

- *Decentralization* - The DS platform should be decentralized and distributed regarding its resources, i.e., computing, storage, and networking for data management. This makes it scalable and near to real-time prototyping in nature. In addition, this platform will be geo-spatially distributed to extend its functionality to target use cases, where this platform serves as a toolchain for data management operations.
- *Semantically Context Awareness* - The DS platform is perceived to be context-aware based on semantics-driven data linkage. This is important to generate knowledge graphs and cross-domain linked information required to build data-driven value chains among stakeholders.
- *Trust and Sovereignty* - This is an important feature in any DS platform that ensures the stakeholder who owns data shall have complete control over their data. This is also needed for General Data Protection Regulation (GDPR) compliance within the EU.
- *Edge-enabled Infrastructure* - DS platform shall be able to realize on-premises near the data sources and with all required relative toolchains available to cater to specific needs for the target use case and related stakeholders. Anyway, in the DS context, the data mostly lies with the generator, and it only expects the data to be searched, indexed, and accumulated on a temporary need basis. Hence, it eliminates the need for expensive cloud-enabled recurring costs and centralized facilities. Therefore, such platforms can be realized with relatively smaller costs.

B. Context Aware Dataspace Model

The system model for our context-aware DS is shown in Figure 4. It is based on the identified requirements and our previous work on the Distributed Edge Network Operations oriented Semantic (i.e. DENOS) model, presented in [7]. It is motivated by the "Onion Architecture" design [28], wherein the key idea is to map the dependencies of the outer layers towards the inner layers and the core, providing a clear separation and segregation of concerns, thus simultaneously improving functional and non-functional concerns. Our proposed architecture has five layers and one main core, explained below from outer to inner direction.

- **Data Source Layer:** This layer represents the source of data that needs to be searched, indexed, queried, etc., by different applications for specific purposes or needs. It stores and manages data to a specific domain or organization and has specific

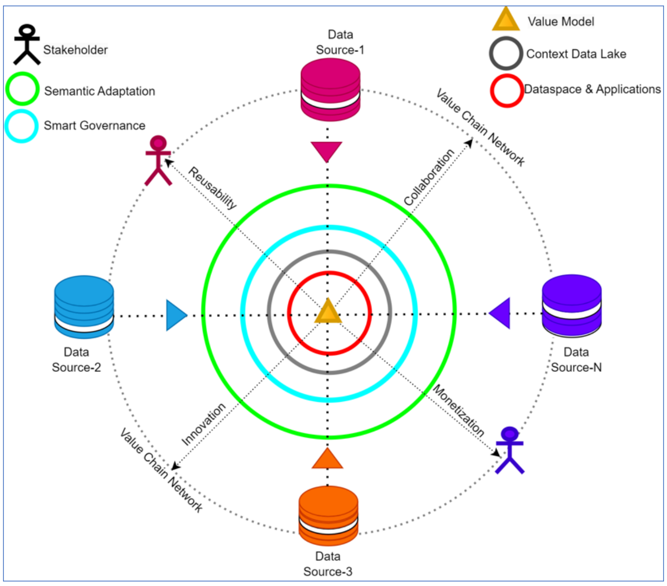

Fig. 4. Context-Aware Dataspace Model.

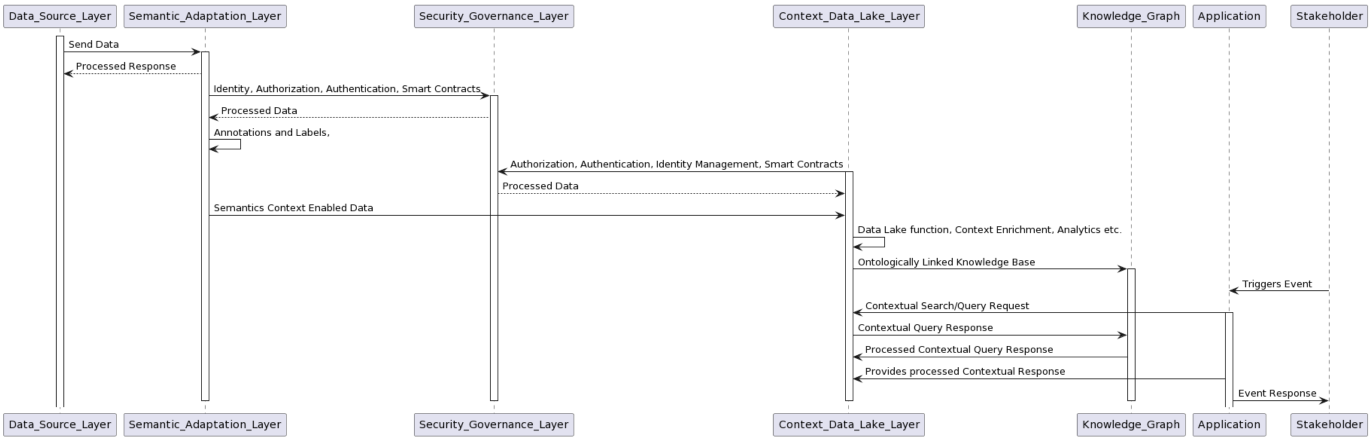

Fig. 5. Sequence Diagram for the Context-Aware Dataspace Operations.

metadata or structure. Different data sources represent different metadata or data models, though they may be semantically identical. Thus, it induces the challenge of heterogeneity and interoperability during the data integration operations.

- **Semantic Adaptation Layer:** To harmonize heterogeneity, this layer provides tools and methods to annotate the incoming data (from the Data Source layer) semantically as per ontology and metadata models. This layer also provides tools to define/reuse relevant ontology and metadata models. Semantic modeling standards like NGSI-LD, RDF, Web Ontology Language (OWL), JSON for Linking Data (JSON-LD), etc. can be used here.
- **Smart Governance Layer:** This layer provides mechanisms to offer identity and access management to maintain trust and sovereignty of the data being operated. This can be achieved using Identity and Role and Attribute access management in a traditional way leveraging standards such as Security Assertion Markup Language (SAML), OpenID Connect (OIDC), OAuth 2.0, System for Cross-domain Identity Management (SCIM), etc., implemented or integrated through DLT/Blockchain-driven smart contracts to have fine-grained granular control [31]. It ensures identity, role, and attribute-based access in a decentralized, transparent, and tamper-resistant manner.
- **Context Data Lake Layer:** This layer represents a specialized data lake offering temporary storage and contextual data management using relevant toolchains. Here, contextual data includes semantically annotated data presenting information at the ontology, domain, and metadata level, providing additional context like metadata, lineage, quality, relationships, origin, etc., for the data to be linked with other domain-level information in different contexts to help machines understand and interpret the data as per the contexts. Further, it facilitates data governance, tracking, discovery, and cataloging efforts, enabling stakeholders to find and utilize the right data for their analytical or operational needs.
- **Application Layer:** This layer provides the DS operations enablement, as per the target use case-driven value context (extraction) needs, over the contextual data in different contexts offered by the contextual data lake.
- **Value Model:** This is the framework's core that triggers different events, such as *Collaboration* for data *Reusability* to *Innovate* new values that can be *Monetized* through building of a *Value Chain Network* among collaborating *Stakeholders* who inspires to derive value out of their *Data Sources*. This drives the value extraction out of the diverse data sources for the given business value context of the use case, leveraging all the upper layers. The business value context can be defined using the relevant business value model, such as St. Gallens Magic Triangle [32] for the given use case.

Functional Flow - Figure 5 illustrates the sequence diagram for the context data lake-centered DS operations. Data comes from the Data Source Layer and enters the Semantic Adaptation Layer, where context annotations and labeling occur using semantic models defined by the domain's ontology. Moreover, before performing adaptation, it requests authorization, authentication, and identity management from the Security Governance Layer based on agreed-upon smart contract-driven policies. Then, the data is ingested inside the Context Data Lake Layer, which holds the data in relevant semantic service context [7] after the Data Lake's pre-configured pipeline operations, such as data/context enrichment, storage, analytics, etc. Thus, the Context Data Lake Layer holds data from multiple sources with multiple semantic contexts and builds a converged knowledge graph for the entire DS model.

C. Deployment Architecture

Multiple reasons motivated us to build DS at the edge. First, the DS is perceived to utilize edge network infrastructure in a coordinated manner, as shown in Figure 6, wherein each edge acts as the organizational entity holding the data with the ownership and providing the relevant semantics context and infrastructure for processing the data at the edge. This offers many advantages, such as the availability of infrastructure by resource pooling across edge networks, which will be a cost-efficient method and allow control of data processing at the edge, thereby raising trust and participatory stake in a multi-stakeholder DS environment. In this context, we intend to emphasize that all participating stakeholders interested in building the DS for mutual benefits can provide the necessary edge network infrastructure required to deploy the proposed DS model. Anyway, saving and optimum utilization of resources at the edge is always the objective of edge computing and the future internet paradigm. Therefore, we have extended an ALC approach [30] to be used in the DS implementation context. ALC provides the flexibility to choose and pick different services from the service catalog and relevant open-source tools, as shown in Figure 7, to develop the pre-configured processing pipeline artifacts to implement the DS layer operations. This way, it helps to choose, select, and deploy only the required services to certain stakeholder or use-case contexts. Thus, saves a lot of computing resources, energy, and cost while addressing the challenges of heterogeneity, integration, and interoperability along with pre-defined processing pipelines and resource requirements. Under the ALC approach, the user selects the packages from the service catalog and generates the relevant artifacts, which can be deployed easily over the edge infrastructure in a distributed manner.

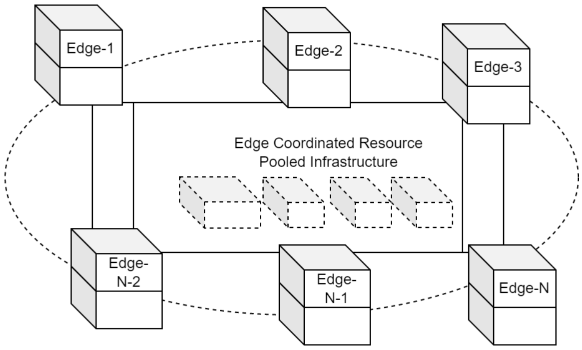

Fig. 6. Edge Coordinated Resource Pooling.

Dataspace Testbed

Layer	Service Catalogue	Realization Tools	Aritifacts for Targeted Use Case: Industrial Projects	R&D Projects	Training & Education
Application	Search & Query Interfaces	SQL/REST/GraphQL	X	X	
	Data Visualization	Grafana		X	X
	Resources Management	Horizon (Openstack)		X	
	Stream Visualization	StreamPipes Connect	X	X	X
Platform	Analytics tool chain	Anaconda/Jupyter	X		X
	Programming modules	Node-Red/Python/NodeJS		X	X
	Semantic Broker	Scorpio NGSI-LD		X	
	Simple Broker	Mosquitto-MQTT Broker	X		X
	Stream Processor	StreamPipes Connect	X		X
	Onotology Modeller	Onotology Modeller		X	X
Virtualization	Virtual Storage	Cinder/Ceph Volumes	X	X	
	Virtual Networks	OpenVswitch/Neutron	X	X	X
	Virtual Routers and Ports	Neturon (Openstack)	X	X	
	Virtual Machines	Libvirt	X	X	X
	Virtual Containers	Docker	X	X	X
Hardware	Connectivity Media	RJ-45, Wifi Adapters	X	X	X
	Router and Switch	Any router and Switch	X	X	X
	Raspberry Pis	Raspberry Pis 4.0		X	X
	Servers/PCs	16GB, 8 cores, 200G HDD, 2 NICs, 1 wifi Adapter	X	X	X
	Internet Connectivity	5G/LTE/Wifi/WLan	X	X	X
Gateway Protocols	MQTT/AMQP	Mosquitto	X		X
	NGSI-LD	Scorpio		X	X
	Modbus/Bluetooth/LoraWan	Modbus/Bluetooth/LoraWan	X		X
	OPC-UA	OPC-UA			X
	HTTP REST	HTTP REST			X
Data Sources	Platforms	AWS, Azure, Private Cloud		X	
	Systems	Data Management Systems	X		X
	Open Data Services/APIs	CKAN, Orion		X	X
	Devices	IoT sensors, Actuators, UAVs	X	X	X
	Gateways	IoT or industrial gateways	X	X	X

Fig. 7. A-La-Carte Approach for Dataspace Model Implementation.

The deployment architecture for the DS platform/testbed is shown in Figure 8. The testbed is developed utilizing on-premises infrastructure and is incrementally scalable. The testbed's infrastructure, system, services, or applications can be scaled without disturbing the existing setup to accommodate the elasticity in the computing and processing demands.

The testbed's infrastructure is provisioned by Kubernetes which is a distributed microservices orchestrator [33]. We have used K3s which is a lightweight distribution of Kubernetes [34]. It supports Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (Saas) models, catering to the diverse needs of stakeholder's use-case in the DS ecosystem. The testbed leveraged Infrastructure-as-a-Code, based on Ansible [33] for bootstrapping of infrastructure. Following this, the PaaS and SaaS are provisioned using the ALC approach, incorporating relevant toolchains like helm charts or Kubernetes templates [33].

The testbed's Platform Layer contains the core implementation of the DS model, encompassing the context data lake functions like data ingestion, authentication, storage, metadata management, and cataloging. The architecture is organized into distinct operational namespaces for resource isolation like (i) *Admin namespace* to manage the infrastructure and resource provisioning, (ii) *Stakeholder-specific namespace* to emulate cross-domain organizational projects for DS with limited access based on predefined roles, along with virtual resource allocation tailored to project needs, (iii) *Common-services namespace* to host shared services like broker, database, NodeRed, and Jupyter, accessible via agreed-upon APIs and permission. The DS testbed is provisioned with various artifacts utilizing Kubernetes/helm-based templates tailored to ALC package selection. These artifacts empower a broad spectrum of services for semantic adaptation and context-aware data lake processing. This encompasses Integration-as-a-Service (e.g., IDS connector) for semantic context-aware data operations, AI-as-a-Service (e.g., StyleGAN) with GPU-enabled edge-instances for machine learning, Database-as-a-Service (e.g., Postgres, and MySQL) for managing diverse types of data, and Programming-as-a-Service (e.g., Node-Red, and WordPress) for custom data processing flow development.

D. Validation

The proposed architecture is validated against the wind turbine use case, presented already in [31]. However, the operations of this use-case have been represented semantically, for the first time, in Figure 9 using RDF standard. This bolt-specific operations semantic model serves as the basis and shows the path to define harmonized cross-domain data models among diverse stakeholders collaborating in wind turbine supply chains in the energy sector. This use case demonstrates the cross-domain digital traceability requirement for bolt, turbine, and related stakeholders that need to deal with diverse events being managed through our local Dataspace testbed. This use-case has been expressed semantically as - *A Service engineer with Name/Employee-ID (Domain-1) performing bolt, coming with Batch-No./ID (Domain-2) coming from supplier with ID, tightening operation at the turbine of turbine operator/manufacturer with turbine ID (Domain-3) at a certain location and time with timestamp*. So, the use-case deals with data from three different domains namely service engineer, turbine operator, and bolt supplier.

The complete functional flow consists of nearby edge to the installed turbine capturing the relevant events (e.g., Service engineer registering for the device at the edge, Bolt batch registration by the turbine manufacturer, turbine/bolt identification via QR code scanning, bolt-supplier mapping registration, etc.) over radio interface (e.g. Bluetooth in our case) in the turbine assembly area or on the field. At the nearby edge, the semantic adaptation (static) functionality is provisioned using the ALC service artifact approach. In this case, the node-red based flow service artifact is provisioned on the edge. This adaptation service receives data over a Bluetooth radio

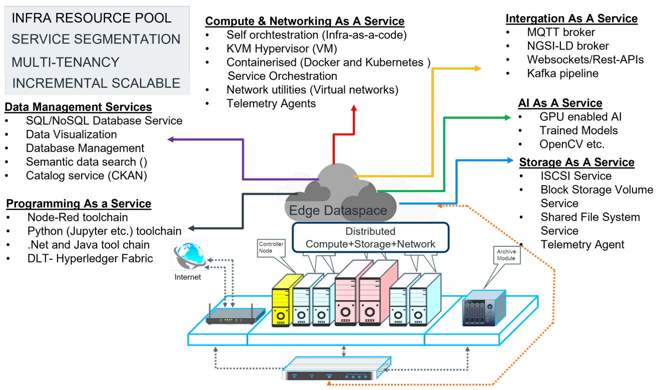

Fig. 8. Deployment Architecture for Dataspace Model.

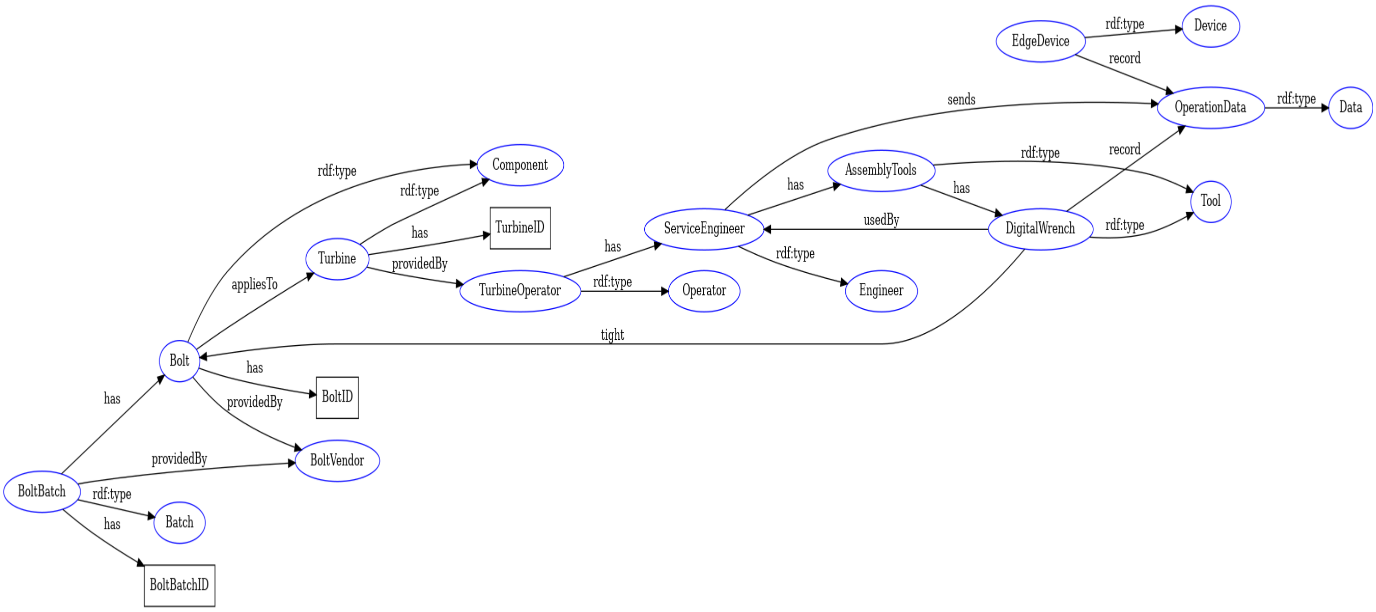

Fig. 9. Digitized Wind Turbine Bolt Semantic Operations Context Model

interface on one hand, and it converges data from different events to create a semantically linked message using the NGSI-LD standard, on the other hand. Afterward, the transformed semantic data is pushed into the permissioned and private Blockchain, implemented using the HyperledgerFabric service artifact, and running at the neighboring edge. This provides us with the smart governance layer based on smart contracts-driven policies validating the pre-registered data model in our case. This can also be used to validate identification, authorization, and authentication through relevant smart contracts in combination with traditional security methods such as identity management or OAUTH2 standards. Finally, the data is processed further for context data lake layer functionality (e.g., StreamPipes, NGSI-LD broker) that allows the building of a knowledge base (based on semantically adapted contexts) and semantic CRUD operations (e.g., SPARQL/NGSI-LD) over data. Different stakeholders can now read this data over semantic contextual interfaces based on their role and permission level agreed upon in smart contracts. To validate this, various cross-domain semantic queries were executed by the stakeholder application, such as - *Fetch bolts from a specific batch ID that impact certain turbines to predict their maintenance requirements or inspection of operational events (e.g., torque value recorded during bolt tightening) for insurance claims under unseen events.*

The average response time results for different operations and their explanation are given in Table I. In addition, this demonstrates the functional validation of the proposed DS model in local and cross-domain contexts. This shows the possibility of a collaborative data-driven value chain development among multiple stakeholders through the proposed model.

The artifacts for this use case consist of frontend (Node-Red, Bluetooth libraries, Web3.js) and backend (REST API, Blockchain Ganache/Hyperledger Fabric) components packaged and provisioned using the ALC approach and Kubernetes orchestrated distributed infrastructure, respectively. The frontend and backend components deployed in different namespaces (representing stakeholder system) over the edge (using two x86 servers-8 core, 16 GB RAM, 80 GB HDD) enabled-DS testbed.

TABLE I
DATA OPERATION AND THEIR RESPONSE TIMES.

Operation Type	**Response Time (ms)**	**Functional Context and Dataspace Model Relevance**
- Registration of Device-turbine or Bolt	800	Stakeholder application registers for turbine or Bolt attributes. - Application, Smart Governance, and Context Data Lake layers are involved.)
Bolt or turbine ID Validation	1200	Service engineer scans the QR code for Turbine and Bolt ID and the relevant event at the edge creates a query to fetch Dataspace from the registered knowledge base. - Data source, Semantic Adaption, Smart Governance, and Context Data lake layers are involved.
Torque Recording	500	Digital wrench is used to tight the bolt, and the relevant torque value is recorded by the nearby edge over Bluetooth and this is then recorded in Blockchain and Application backend both. - Data source, Semantic Adaption, Smart Governance, and Context Data lake and application layers are involved.
Read Turbine, Bolt, or Log entry	600	Application interface reading the Dataspace backend for relevant event data. - Application, Smart Governance, and Context Data Lake are involved.

V. CONCLUSIONS AND OUTLOOK

This paper has introduced the motivation for developing a DS platform at Edge and its realization being presented for the first time in literature, along with the background and relevant work in this area. This study identifies the requirements for edge-enabled DS based on discussions with stakeholders dealing with different data integration, reusability needs, and desire for integrated value-chain development. As a result, a novel context-aware DS model with semantic capabilities is proposed and prototyped in a lab environment. In addition, the deployment is supported by the edge-oriented resources pool infrastructure and orchestrated following the extended ALC approach based on predefined service catalog artifacts. The proposed DS model is also validated against identified requirements following a wind turbine use case. As an outlook, we would like to detail this model further for each layer with concrete implementation for diverse cross-domain use cases. Finally, this study contributes knowledge on how context-aware DS ecosystems for data integration can be realized in local or regional contexts at a small scale by exploiting relevant resources at the edge in real-world scenarios. In addition, this study advances the knowledge on the use of semantic adaptation, smart governance, and context-aware data lake for enhancing the efficiency of cross-domain data management operations and value chain development. Thus, adding value in the context of evolving industry5.0 ecosystems and upcoming technologies, such as 6G, in the future internet landscape.

References

[1] G. Solmaz, F. Cirillo, J. Fürst, T. Jacobs, M. Bauer, E. Kovacs, J. R. Santana, and L. Sánchez, "Enabling data spaces: Existing developments and challenges," in *Proceedings of the 1st International Workshop on Data Economy*, 2022, pp. 42–48.

[2] J. Baloup, E. Bayamlıoğlu, A. Benmayor, C. Ducuing, L. Dutkiewicz, T. Lalova-Spinks, Y. Miadzvetskaya, and B. Peeters, "White paper on the Data Governance Act," 2021.

[3] P. Singh, A. U. Haq, Nidhi, and M. Beliatis, "Meta standard requirements for harmonizing dataspace integration at the edge," in *2023 IEEE Conference on Standards for Communications and Networking (CSCN)*. IEEE, 2023, pp. 130–135.

[4] S. R. Jeffery, M. J. Franklin, and A. Y. Halevy, "Pay-as-you-go user feedback for dataspace systems," in *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, 2008, pp. 847–860.

[5] H. Ning and T. Wang, "Constructing a dataspace based on metadata and ontology for complicated scientific data management," in *2007 2nd International Conference on Pervasive Computing and Applications*. IEEE, 2007, pp. 512–514a.

[6] E. Curry, W. Derguech, S. Hasan, C. Kouroupetroglou, and U. ul Hassan, "A real-time linked dataspace for the internet of things: enabling "pay-as-you-go" data management in smart environments," *Future Generation Computer Systems*, vol. 90, pp. 405–422, 2019.

[7] P. Singh, M. J. Beliatis, and M. Presser, "Enabling edge-driven dataspace integration through convergence of distributed technologies," *Internet of Things*, vol. 25, p. 101087, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2542660524000295

[8] E. Curry, S. Scerri, and T. Tuikka, *Data Spaces: Design, Deployment and Future Directions*. Springer Nature, 2022.

[9] T. Heath and C. Bizer, *Linked data: Evolving the web into a global data space*. Springer Nature, 2022.

[10] Q. Duan, S. Wang, and N. Ansari, "Convergence of networking and cloud/edge computing: Status, challenges, and opportunities," *IEEE Network*, vol. 34, no. 6, pp. 148–155, 2020.

[11] B. Farahani, F. Firouzi, and M. Luecking, "The convergence of iot and distributed ledger technologies (dlt): Opportunities, challenges, and solutions," *Journal of Network and Computer Applications*, vol. 177, p. 102936, 2021.

[12] H. Lee, B. Lee, H. Yang, J. Kim, S. Kim, W. Shin, B. Shim, and H. V. Poor, "Towards 6g hyper-connectivity: Vision, challenges, and key enabling technologies," *Journal of Communications and Networks*, 2023.

[13] B. Otto, "GAIX-X and IDS. Position Paper, Version 1.0 01," 2021. [Online]. Available: https://doi.org/10.5281/zenodo.5675897

[14] P. Zehnder, P. Wiener, T. Straub, and D. Riemer, "Streampipes connect: semantics-based edge adapters for the iiot," in *The Semantic Web: 17th International Conference, ESWC 2020, Heraklion, Crete, Greece, May 31–June 4, 2020, Proceedings 17*. Springer, 2020, pp. 665–680.

[15] A. Hutterer and B. Krumay, "Integrating heterogeneous data in dataspaces-a systematic mapping study," *Pacific Asia Conference on Information Systems*, 2022.

[16] J. Hernandez, L. McKenna, and R. Brennan, "Tikd: A trusted integrated knowledge dataspace for sensitive healthcare data sharing," in *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE, 2021, pp. 1855–1860.

[17] A. Seidel, K. Wenzel, A. Hänel, U. Teicher, A. Weiß, U. Schäfer, S. Ihlenfeldt, H. Eisenmann, and H. Ernst, "Towards a seamless data cycle for space components: considerations from the growing european future digital ecosystem gaia-x," *CEAS Space Journal*, pp. 1–15, 2023.

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

[19] E. Commission, "Data act," 2022. [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/data-act

[20] A. Paulus, A. Pomp, and T. Meisen, "The plasma framework: Laying the path to domain-specific semantics in dataspaces," in *Companion Proceedings of the ACM Web Conference 2023*, 2023, pp. 1474–1479.

[21] R. Angles, H. Thakkar, and D. Tomaszuk, "Rdf and property graphs interoperability: Status and issues." *AMW*, vol. 2369, 2019.

[22] F. Martella, V. Lukaj, M. Fazio, A. Celesti, and M. Villari, "On-demand and automatic deployment of microservice at the edge based on ngsi-ld," in *2023 31st Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)*. IEEE, 2023, pp. 314–320.

[23] E. CIM, "Ngsi-ld cim 009," ETSI, Tech. Rep., 2019, times cited: 3. [Online]. Available: https://www.etsi.org/deliver/etsigs/CIM.

[24] J. Möller, D. Jankowski, A. Lamm, and A. Hahn, "Data management architecture for service-oriented maritime testbeds," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 3, pp. 631–649, 2022.

[25] J. M. Bryson, "What to do when stakeholders matter: stakeholder identification and analysis techniques," *Public management review*, vol. 6, no. 1, pp. 21–53, 2004.

[26] V. Niranjan, S. Anand, and K. Kunti, "Shared data services: An architectural approach," in *IEEE International Conference on Web Services (ICWS'05)*. IEEE, 2005.

[27] S. Bader *et al.*, "The international data spaces information model–an ontology for sovereign exchange of digital content," in *International Semantic Web Conference*. Springer, 2020, pp. 176–192.

[28] M. E. Khalil, K. Ghani, and W. Khalil, "Onion architecture: a new approach for xaas (every-thing-as-a service) based virtual collaborations," in *2016 13th Learning and Technology Conference (L&T)*. IEEE, 2016, pp. 1–7.

[29] I. K. Aksakalli, T. C, A. B. C, and B. T, "Deployment and communication patterns in microservice architectures: A systematic literature review," *Journal of Systems and Software*, vol. 180, p. 111014, 2021.

[30] J. Sibold, "Learning" a la carte": A theory-based tool for maximizing student engagement." *Journal of College Teaching & Learning*, vol. 13, no. 2, pp. 79–84, 2016.

[31] P. Singh, K. Holm, M. J. Beliatis, A. Ionita, M. Presser, P. Wolfgang, and R. C. Goduscheit, "Blockchain for economy of scale in wind industry: A demo case," in *Global IoT Summit*. Springer, 2022, pp. 175–186.

[32] O. Gassmann, K. Frankenberger, and M. Csik, "The st. gallen business model navigator," *Int. J. Prod. Dev*, vol. 18, pp. 249–273, 2013.

[33] L. Berton, "Ansible for k8s management," in *Ansible for Kubernetes by Example*. Springer, 2023, pp. 201–237.

[34] Y. Hao, "Edge computing on low availability devices with k3s in a smart home iot system," Ph.D. dissertation, The Cooper Union for the Advancement of Science and Art, 2022.