DATA-DRIVEN IOT ECOSYSTEM FOR CROSS BUSINESS GROWTH: AN INSPIRATION FUTURE INTERNET MODEL WITH DATASPACE AT THE EDGE

AUTHORS: PARWINDER SINGH, NIDHI, MICHAIL J. BELIATIS, MIRKO PRESSER
EMAIL: PARWINDER@BTECH.AU.DK

ABOUT ME

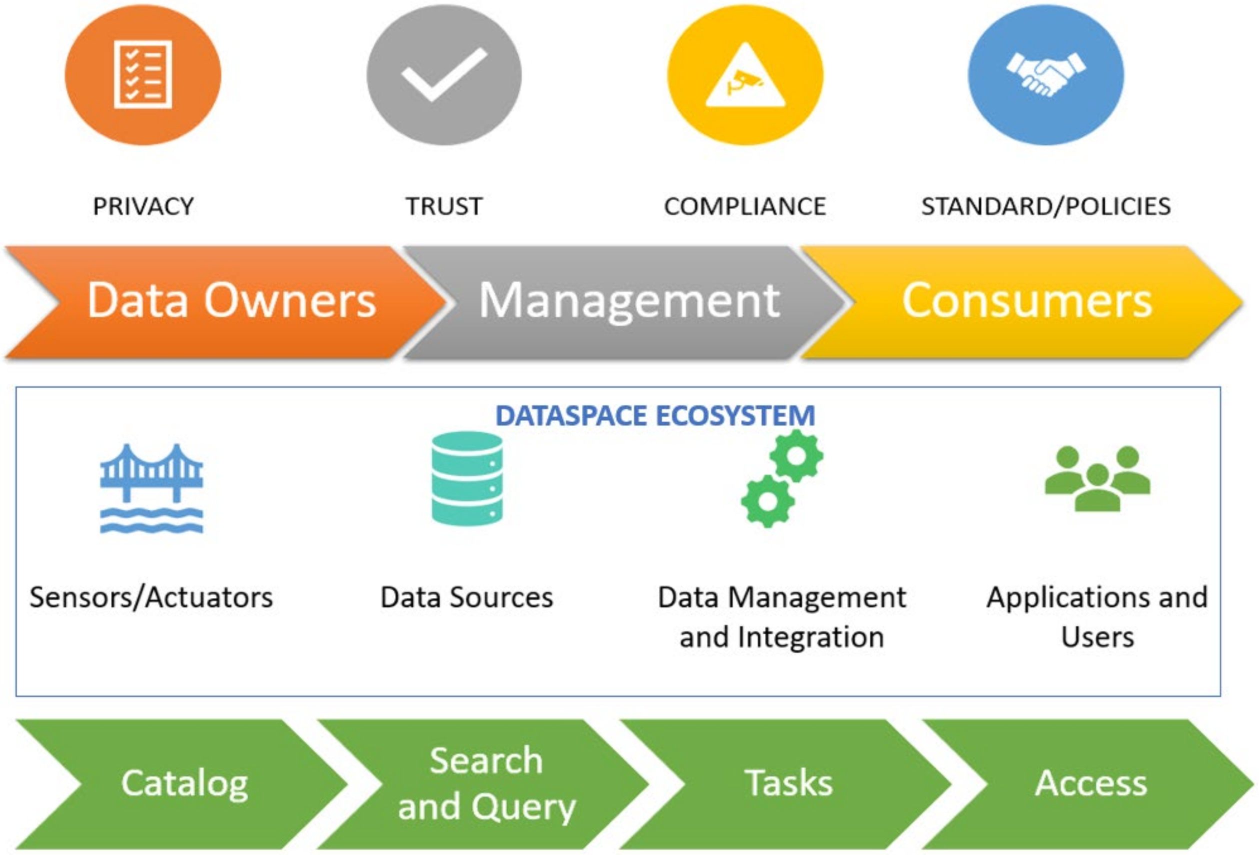
- 2007: **Bachelor in Electronics and Telecommunication**, IETE, New Delhi, India
- 2007-2008: Worked as **Network and system administrator**, Genius Informatics, Delhi, India.
- 2009-2011: **MSc Electronics Engg**. (spec. Comm. Systems), Hochschule Bremen, Germany.
- 2012-2014: **Sr. Software Engineer**, Aricent Technologies in TCP/IP, Security, Telecom area.
- 2014-2017: **Technical Leader Engineering**, Aricent Technologies in SDN, NFV/MANO, Cloud.
- 2017-2018: **Sr. Technical Leader**, NEC Technologies in Cloud Computing design & development.
- 2018-2020: **System Architect**, NEC Technologies in IoT Smart City product development.
- 2020-2021: **Research Assistant**, Aarhus University, Lead developer in EU IoT Crawler Project.
- 2022-Present: **Ph.D. student (DBD, Aarhus University):** Topic- Convergence of technologies at edge for industry4.0 & beyond
- Around **30+ industry projects** delivered successfully in the career of 10 years span.
- Around 10+ research papers so far and contributed to many open source communities in project/product management, development, and quality management like FIWARE, Cloud computing, and IoT along with many R&D prototypes development.

OVERVIEW

- Data -> **bloodline** for a business to **grow, compete, and sustain.**
- Traditional data processing **confines value** within organizational boundaries.
- Opportunity for the **cross-organizational services-driven value-chain** system with **win-win** business model development.
- **Dataspace (DS)** offers an ecosystem for data sharing and reuse with ownership, governance, and control.
- Proposed **Dataspace-at-the-Edge model** for future internet and industry 5.0-oriented **cross-boundary** data integration solution in **local or regional contexts.**
  - Each **organization** reflects **an edge** that supports **DS requirements** and offers **edge-oriented advantages (latency, bandwidth).**
  - **Context enablement** of the data through **semantics-linked data lake and semantic adaption**.

MOTIVATION

- **Pathways** to build **data-driven** and **cross-organization value chains**, facilitating diverse **services enablement** and **monetization** opportunities.
- **Challenges**: lack of **context** and **quality**, **heterogeneity**, **dynamic deployment**, **interoperability**, **security**, **trust,** etc.
- The central **Research Question** (RQ): "**How to build DS in the local context for developing data-driven cross-domain service value-chain enablement?**"



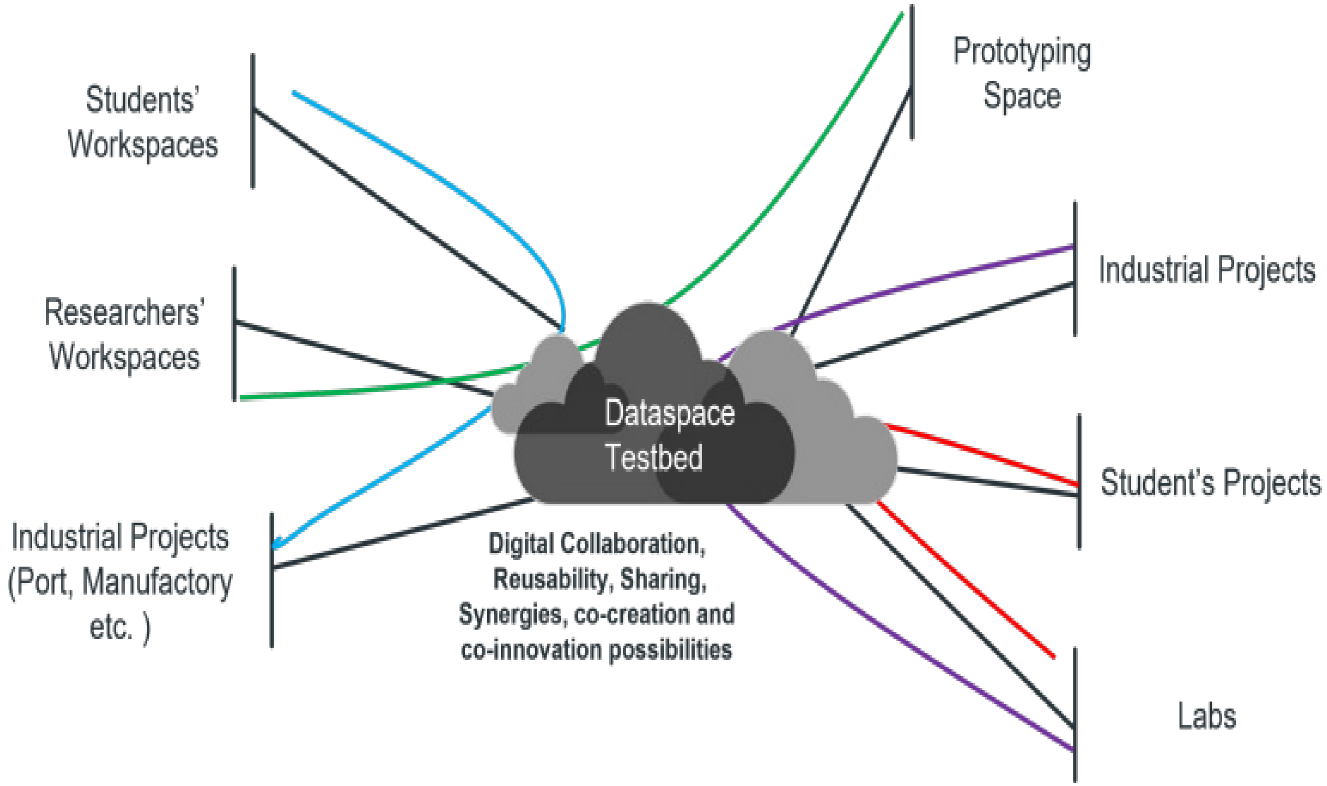
OUR CONTRIBUTION

- Expand **Dataspace applicability** in **local or regional contexts.**
- **Context-aware DS at the Edge** as a potential solution to allow **data and services** to be **shared, reused, exchanged, and integrated** with **governance and control** across domains**.**
- **Semantics enablement** of **the data and associated services** based on a **context-linked data lake and semantic adaption capabilities.**
- A novel **service artifact methodology (A-la-Carte - ALC)**, consisting of a service catalog and relevant toolchain, is also introduced to realize such a DS model efficiently over a distributed edge network.

LITERATURE FINDINGS

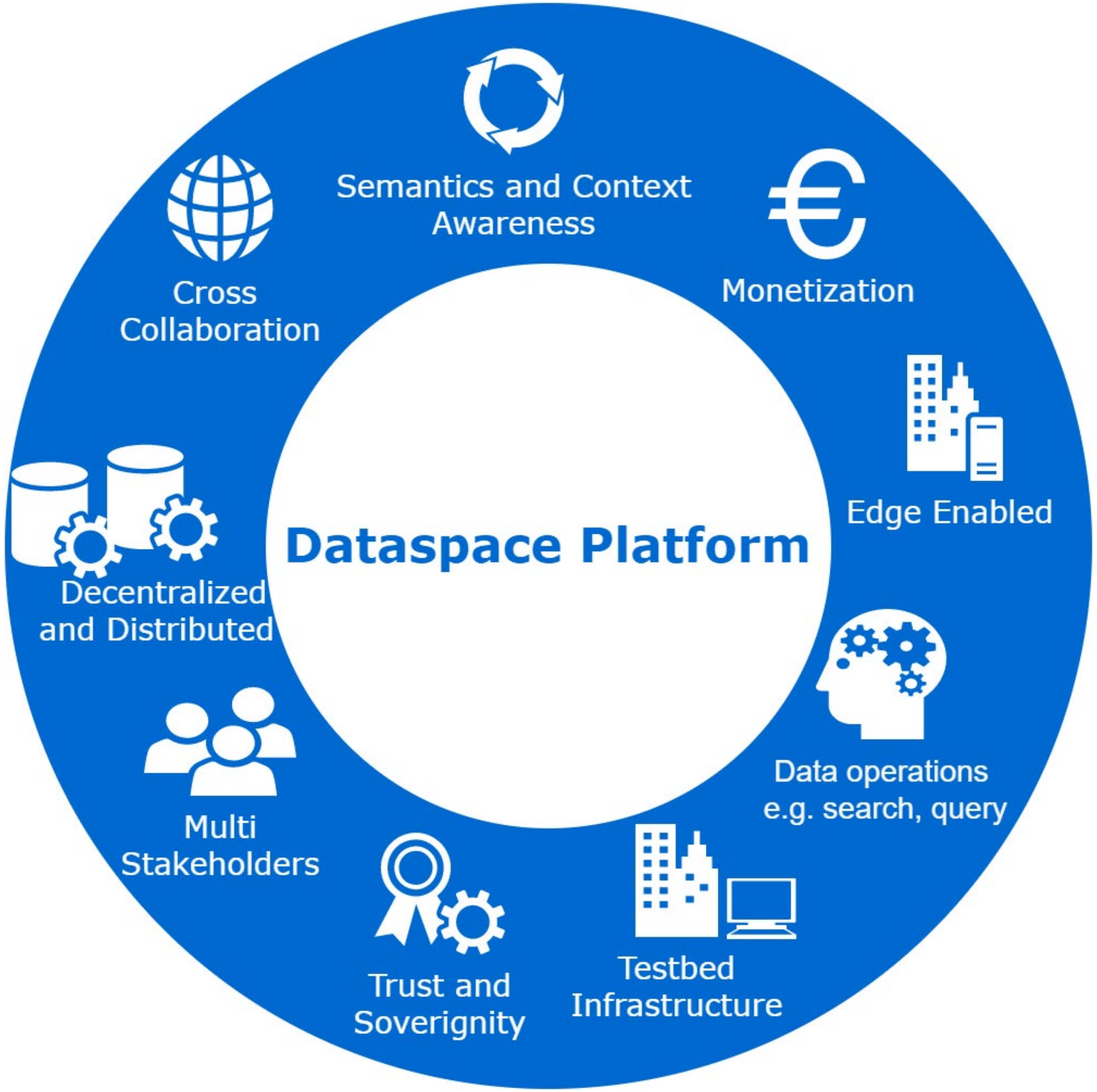
- DS ecosystem **breaks down data silos** and can promote **cross-domain data sharing** with **contextual semantics**.
- **Realizing** DS needs to address **heterogeneity.**
  - Leverage **semantics** wherein **ontologies** represent **machine-readable conceptualization**
    - **Metadata and ontology** are essential for developing **semantic information** to map **business-level domain information** into relevant **technical-leve**l information.
- **Initiatives** like the **International Data Space (IDS) and GAIA-X** in the EU have outlined architectural frameworks for DS ecosystems.
- A wide range of **tools** are available for **distributed edge-driven semantic adaptation** e.g., **StreamPipes Connect, Plasma, RDF, NGSI-LD,** etc.
- DS **doesn't control** the data sources; administration falls under the individuals or their relevant management systems.
- **Limited literature for building DS at edge** in local context for cross-domain ops.

METHODOLOGY



- ❑ Requirements specified in the local context and for different stakeholders.
- ❑ Data and associated operations are common grounds.
- ❑ Vision to extract information in diverse semantic contexts.
- ❑ Sovereignty and Governance to be ensured.
- ❑ Readily available toolchain to perform certain semantic operations over data in a context-aware manner.
- ❑ DS testbed development.

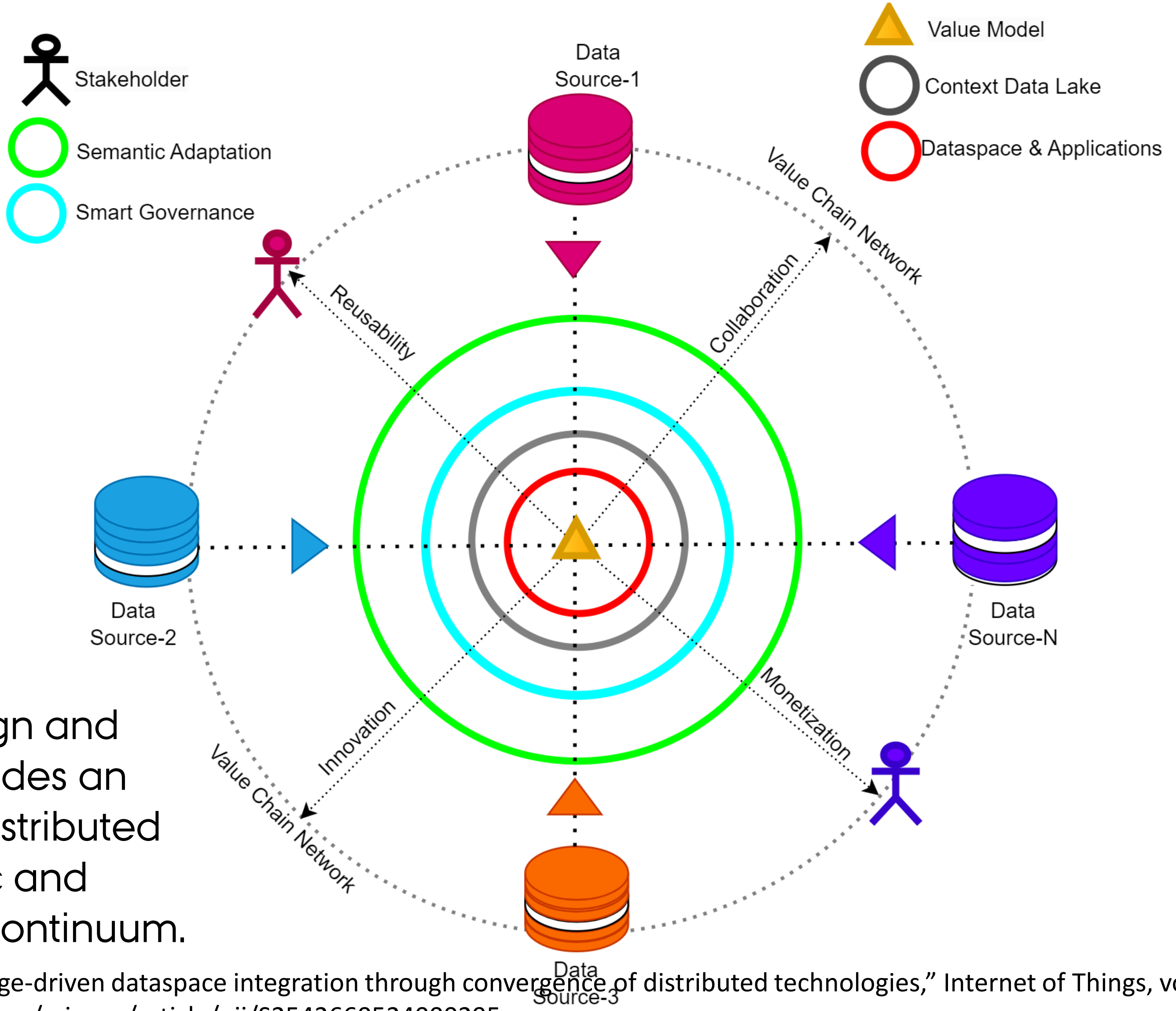
REQUIREMENTS FOR EDGE-DS



CONTEXT-AWARE DATASPACE MODEL

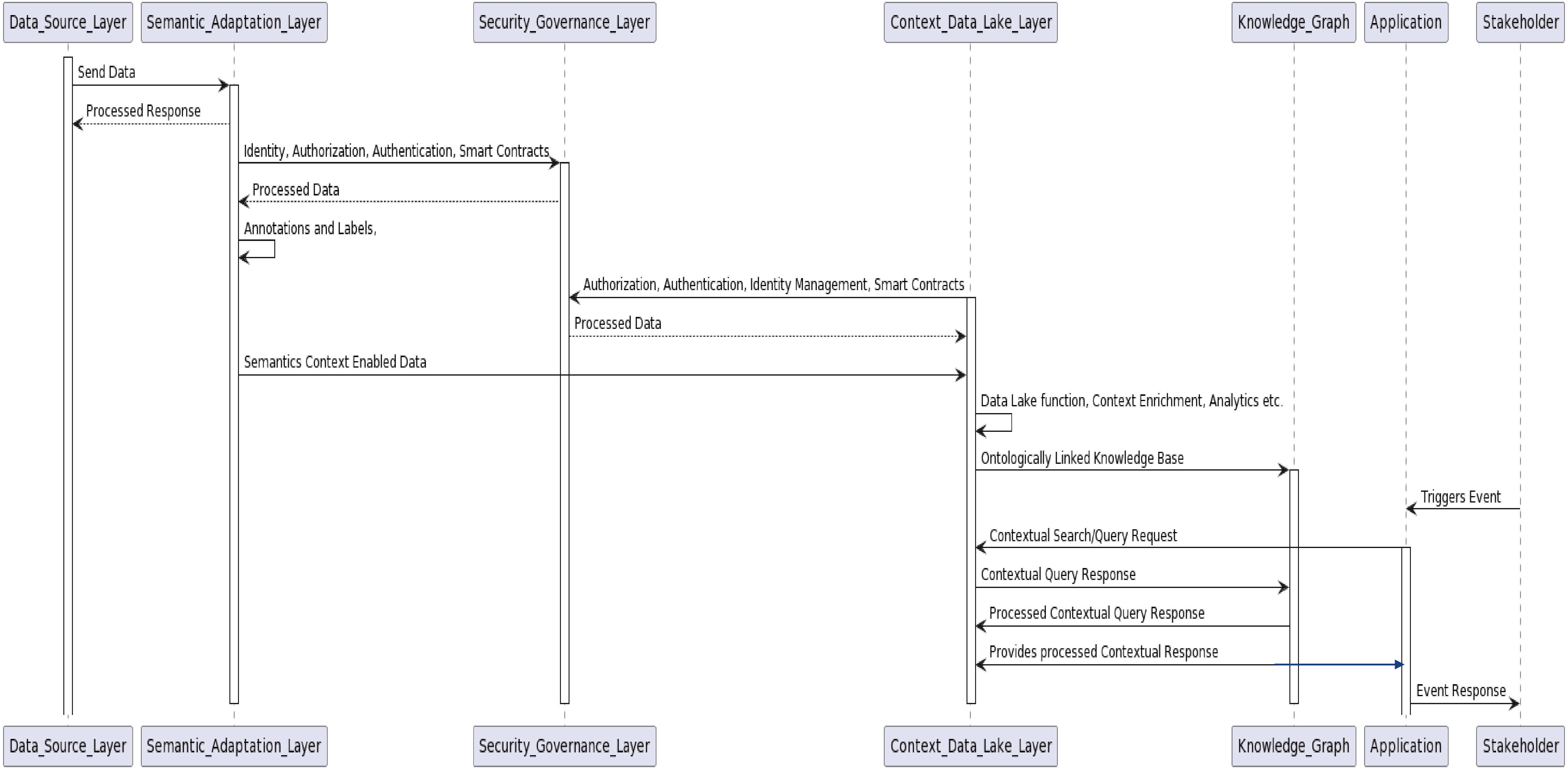
*This framework triggers different events, such as **Collaboration** for **data Reusability** to **Innovate** new values that can be **Monetized** through building of a **Value Chain Network** among collaborating **Stakeholders** who are inspired to derive **value** from the converged **Data Sources**.*

Motivated by the "Onion Architecture" design and extension of the work done in [1] that provides an architectural framework for dynamic and distributed Edge Network operations through semantic and convergence layers to the traditional CED continuum.



[1] 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

SEQUENCE FLOW FOR THE CONTEXT-AWARE DATASPACE OPERATIONS



DEPLOYMENT ARCHITECTURE

INFRA RESOURCE POOL

SERVICE SEGMENTATION

MULTI-TENANCY

INCREMENTAL SCALABLE

Compute & Networking As A Service

- Self orchtestration (Infra-as-a-code)
- KVM Hypervisor (VM)
- Containerised (Docker and Kubernetes) Service Orchestration
- Network utilities (Virtual networks)
- Telemetry Agents

Intergation As A Service

- MQTT broker
- NGSI-LD broker
- Websockets/Rest-APIs
- Kafka pipeline

Data Management Services

- SQL/NoSQL Database Service
- Data Visualization
- Database Management
- Semantic data search ()
- Catalog service (CKAN)

AI As A Service

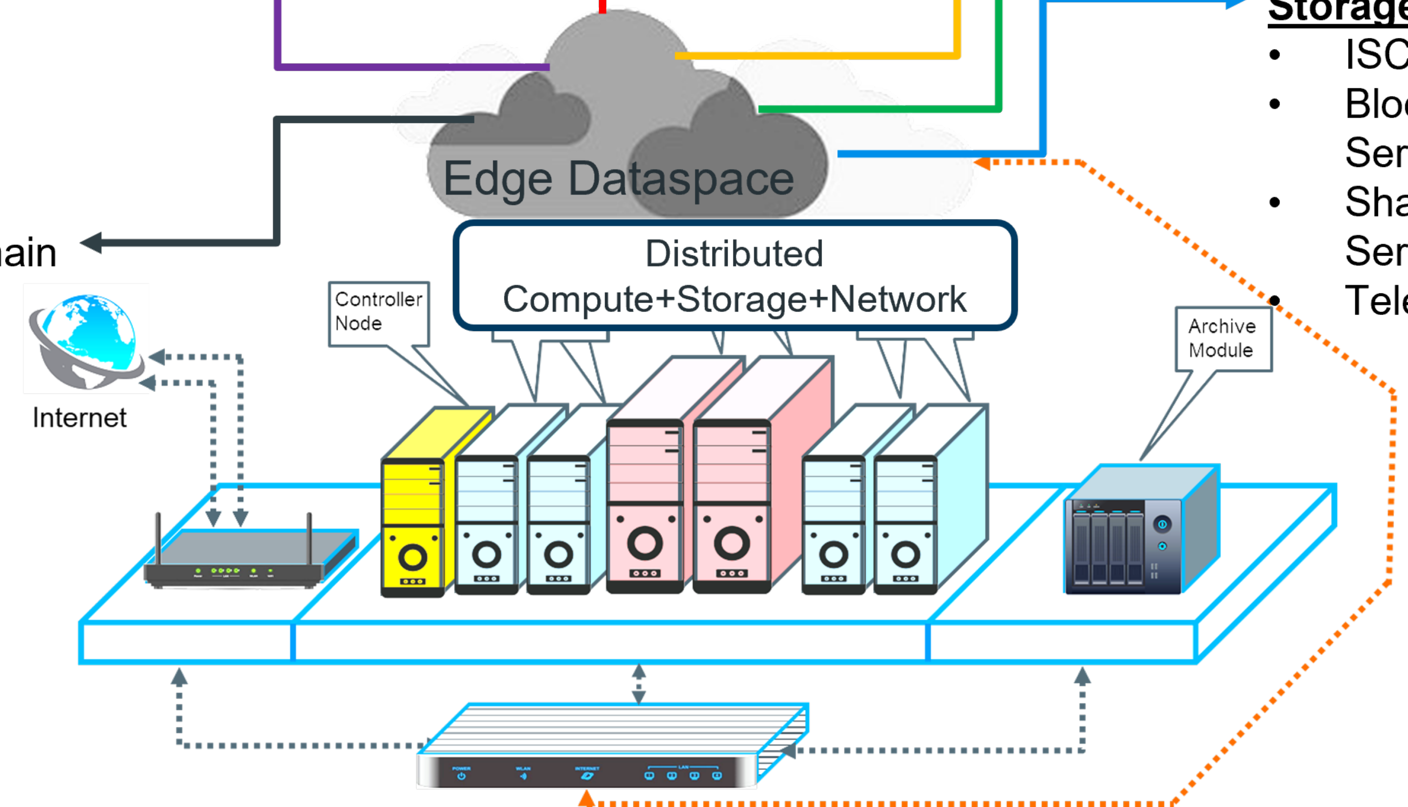
- GPU enabled AI
- Trained Models
- OpenCV etc.

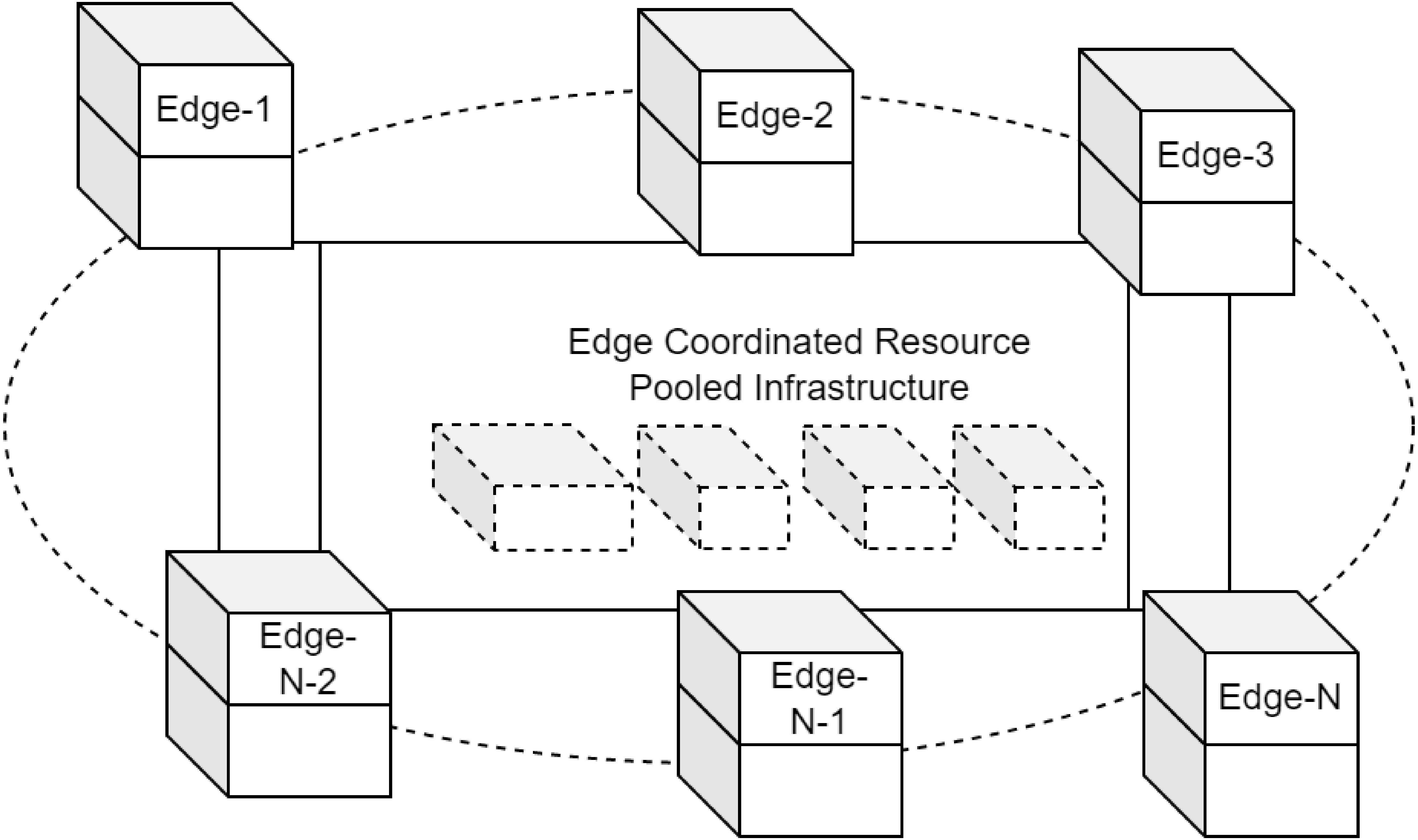
Storage As A Service

- ISCSI Service
- Block Storage Volume Service
- Shared File System Service
- Telemetry Agent

Programming As a Service

- Node-Red toolchain
- Python (Jupyter etc.) toolchain
- .Net and Java tool chain
- DLT- Hyperledger Fabric




Edge-1
Edge-2
Edge-3
Edge Coordinated Resource
Pooled Infrastructure
Edge-
N-2
Edge-
N-1
Edge-N

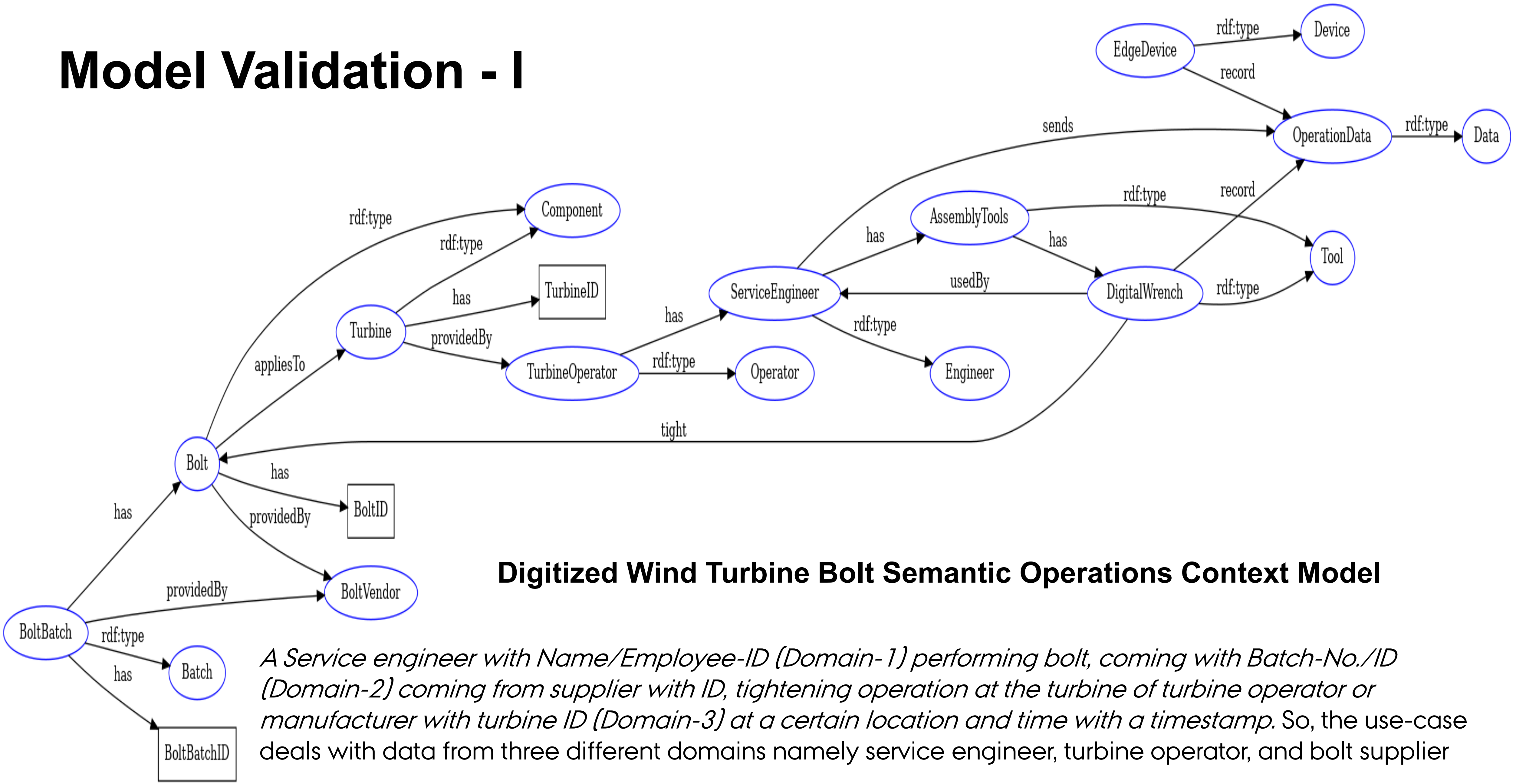
A-LA-CARTE APPROACH [1] FOR EDGE-DATASPACE MODEL IMPLEMENTATION

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

[1] 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.

Model Validation - I



Digitized Wind Turbine Bolt Semantic Operations Context Model

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 or manufacturer with turbine ID (Domain-3) at a certain location and time with a timestamp. So, the use-case deals with data from three different domains namely service engineer, turbine operator, and bolt supplier

Model Validation - II

- **Frontend components:**
  - Node-Red
  - Bluetooth libraries
  - Web3.js
- **Backend components:**
  - REST API
  - Blockchain (Ganache/Hyperledger Fabric)
- **Deployment approach:**
  - ALC approach
- **Infrastructure:**
  - Kubernetes (K3s) orchestrated distributed edge infrastructure
  - Utilizing two x86 servers:
    - 8 core CPU
    - 16 GB RAM
    - 80 GB HDD
- **Deployment setup:**
  - Frontend and backend components deployed in different namespaces
  - Namespaces represent stakeholder systems

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.

CONCLUSION

- **Motivation**: Cross-organization data integration necessitates a DS platform at the Edge.
- **Focus**: Requirements for edge-enabled DS, emphasizing data integration, reusability needs, and integrated value-chain development.
- **Innovation**: Proposed a novel context-aware DS model with semantic capabilities.
- **Implementation**: Prototype of edge-enabled DS developed in a lab environment.
- **Deployment Strategy**: Dynamic deployment supported by the Aa-la-Carte (ALC) approach.
- **Validation**: Model validated against cross-domain semantic wind supply chain operations, found functionally compliant.
- **Contribution**: Expands Knowledge Base for realizing DS in local or regional contexts
- **Impact**: Adds value in the context of evolving industry 5.0 ecosystems and future technologies like 6G in the internet landscape.

THANKS & QUESTIONS?

AARHUS
BSS

DEPARTMENT OF BUSINESS DEVELOPMENT AND TECHNOLOGY

AARHUS UNIVERSITY