On the Need of International Cross-Data Space Interworking: An EU–Japan Case Study

Juan Ramón Santana and Luis Sánchez, *Universidad de Cantabria, 39005, Santander, Spain*
Martin Bauer, Benjamin Hebgen, and Ernö Kovacs, *NEC Laboratories Europe, 69115, Heidelberg, Germany*
Satsuki Hamaguchi and Yuriko Nomura, *NEC Corporation, Tokyo, 108-8001, Japan*

Data spaces have emerged as pivotal elements promoting data-driven applications and driving the growth of the data economy. Opposite to traditional data exchange, where trustworthiness relies on a central entity acting as a data transactions' moderator, the decentralization introduced by data spaces overcomes the barriers for a worldwide market of data economy, ensuring the self-sovereignty for data owners. However, existing solutions for the deployment of data spaces jeopardize their adoption, isolating the data economy in different regions. This article delves into cross-data space interworking between International Data Spaces (IDS) and Connector Architecture for Decentralized Data Exchange (CADDE) architectures as they are the references, in Europe and Japan, respectively, for the creation of this kind of data sharing ecosystems. The article describes a CO_2 footprint assessment case study—the first of its kind as far as we know—enabling cross-domain data exchange between different data spaces and discusses the pilot results.

Data are a key driver for future innovation and economic growth in nearly all activities. Novel artificial intelligence (AI) applications are already improving existing services by increasing their efficiency in providing added value to customers. Data ecosystems are playing an essential role in delivering the vast amount of data required for such purposes. In fact, this is already an important market, having a real impact in different countries. Authors in Micheletti et al.[1] describe the rapid growth of the European data economy—from €443 billion (3.6% of GDP) in 2021 to an estimated €787 billion (5.3% of GDP) by 2030.

In this context, data spaces emerge as an abstraction for decentralized data management, establishing a common set of data governance guidelines. Stakeholders can adhere to these guidelines in a decentralized ecosystem, enabling peer-to-peer data exchange between parties without intermediaries. Data spaces focus on data sovereignty, where data owners keep control of their data throughout the whole data exchange process.

The irruption of solutions scattered across different geographic regions, such as the International Data Spaces (IDS)[2] or Gaia-X,[3] in Europe, and the Connector Architecture for Decentralized Data Exchange (CADDE)[4] in Japan, limits the scope of the data economy and prevents data exchange on a global scale. In fact, policy-makers have concerns about becoming dependent on foreign technology and losing control over the data governance policies they want to implement.

Therefore, the coexistence of heterogeneous solutions has to be tackled to open up the world to a global economy based on data. Data space interworking enables the introduction of novel use cases where data sources are potentially scattered across the world. Hence, requiring the interconnection of data spaces across technological and legislative boundaries.

Digital Object Identifier 10.1109/MITP.2025.3545214
Date of current version 15 October 2025.

The article focuses on cross-data space interworking between CADDE and IDS. Throughout this article we present our work on the interworking between these two different systems, reviewing the current state of data space standards globally. Moreover, the article describes a proof-of-concept (PoC) scenario through which we have demonstrated the feasibility of the solutions that we have proposed to support the interworking between two data spaces deployed using the abovementioned two reference architectures. Hence, there are two main contributions presented in this article. First, it outlines the key challenges to be overcome to address the creation of global data spaces (in contrast with current region-wise data spaces) as well as the solutions proposed to address them. Second, it validates our approach by implementing and testing a PoC use case that is employing our proposed solution for data space interworking.

The article is structured as follows. The "Data Space Ecosystem Overview" section reviews the current data space ecosystem in Europe and Japan. The "Data Space Interworking: Proof-of-Concept Architecture" section outlines the architecture for an interworking solution and presents its implementation in two data spaces from both regions. The "CO_2 Footprint Monitor" section sketches a case study that leverages the cross-data space interworking enabled by our solution. Finally, the "Discussion and Lessons Learned" section discusses the takeaways extracted from our experience, while the "Conclusion" section concludes the article.

DATA SPACE ECOSYSTEM OVERVIEW

The current technology landscape of data spaces is evolving at a dazzling pace. The most mature solution in terms of existing technical developments relates to the IDS. Second generation of data space technologies like Gaia-X are further fostering the decentralized nature of data spaces. Research is ongoing which, eventually, will lead to further generations. In the following, we summarize the key aspects of the leading technologies that are related with our work.

IDS

The IDS architecture provides a reference framework for decentralized data exchange, including common domain-agnostic data models and protocols.[5] The architecture is centered around the notion of a connector, supported by a set of centralized components provided by a data space facilitator. Connectors are required to carry out any interaction in a data space, as they are responsible for establishing secure and trustworthy communication between participants. Before data exchange takes place, participants' connectors agree on a mutually binding contract establishing the conditions under which data can be accessed and used. IDS requires any component interacting in a data space to be certified to guarantee conformance to the processes defined in the architecture, so that agreements can be technically enforced.

At present, the development of a new standard, the Dataspace Protocol,[6] which establishes the control mechanisms for the data exchange, but not for the data transfer itself, focuses the efforts within the IDS reference architecture.

GAIA-X

Gaia-X specifies a framework for creating a federated and secure data infrastructure based on a common trust framework.[7] In a Gaia-X data space, all data are kept by the data provider, and it is only exchanged after negotiating an agreement with the data consumer that specifies how the data can be used.

Gaia-X defines a set of federation services with a focus on the compliance services, on which the trust framework is built. These services include the functionality needed for describing, managing, and discovering data exchange services, as well as negotiating the exchange of data. However, the actual exchange of data is out of scope, and different data transfer technologies can be used.

To establish trust, product offerings, including data and services, have self-descriptions supported by verifiable credentials, which are issued by trust anchors endorsed by Gaia-X. For this purpose, self-sovereign identities (SSI) are the basis to support digital identities.

CADDE

CADDE is a cross-domain data exchange platform for discovery and cross-domain data sharing. The core element is a network of connector modules enabling data providers and consumers to participate in a CADDE data space. CADDE is planned to be used by the Japanese nation-wide cross-domain data exchange platform DATA-EX. With the support of CADDE, DATA-EX provides the following:

- unique identification of data items
- secure access to the data space only to authorized members
- connectors enabling publishing, discovery, and using data under access control policies

- data catalogues managing metadata descriptions and data access details
- history management controlling relationships between data items, logging all data exchanges.

FIWARE

FIWARE is an open source ecosystem providing components for building smart solutions. At the core of FIWARE are context brokers (CBs) that enable the exchange of data, connecting various stakeholders.[8] CBs implement the Next Generation Service Interface-Linked Data (NGSI-LD) application programming interface (API), which is standardized by ETSI.[9] NGSI-LD models context (i.e., the world, in general) as typed entities that have properties and relationships to other entities, following a semantic approach.

DATA SPACE INTERWORKING: PROOF-OF-CONCEPT ARCHITECTURE

To demonstrate the potential of intercontinental data space interworking, we propose a PoC architecture to enable the integration of two data spaces based on CADDE and IDS.

A first approach would be to simply consider a proxy (i.e., brokering the requests toward data providers), where a single component, such as a dedicated connector, conceals the participants from the other data space. However, apart from potential scalability issues, this solution limits a thorough auditability of transactions between participants, and this dedicated connector would need to be certified to operate in both data spaces, leading to potential conflicts. It also implies that data space providers would not control their own policies (built-in regulations and laws), as they would have to rely on the intermediary (i.e., the dedicated connector).

Considering these restrictions, we are proposing the data space peering concept (cf. Figure 1) for the interworking architecture, where a data space interworking unit (DS-IWU) enables data exchange between two participants from different data spaces. It can be deployed by any participant involved in the data exchange but is typically done by providers interested in offering their data. Once the participant obtains the credentials for both data spaces, the identities associated with the deployed DS-IWU are seamlessly linked across them, ensuring correct auditing of transactions and the enforcement of access and usage control rules. Moreover, the DS-IWU provides a method for policy translation and ensures correct data handling within the transaction.

The DS-IWU comprises an IDS and a CADDE connector, as mandatory elements on both sides, and one intermediate module to perform the corresponding conversions, including identity and policy management. It also implements a catalog exchange method to provide data discovery mechanisms for both data spaces, through the conversion of metadata descriptions in their respective data discovery solutions.

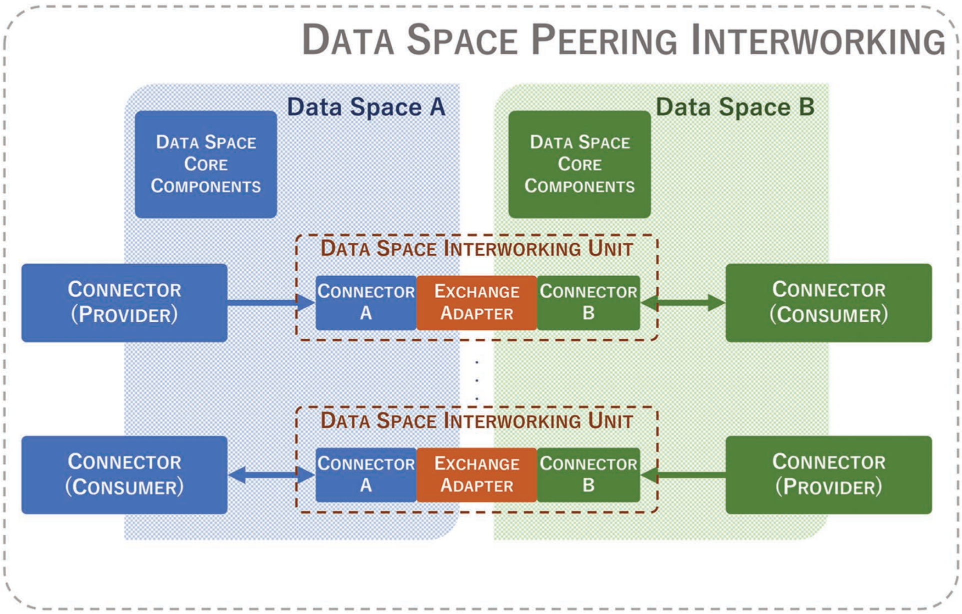

FIGURE 1. Data space peering interworking architecture.

It is important to highlight that the pilot that has been developed focuses on IDS-based and CADDE-based data spaces interworking, but for the DS-IWU to enable interworking with other data spaces, the corresponding connector should be plugged in and the exchange adapter configured to adapt to the required conversions.

Data model heterogeneity was supported leveraging the NGSI-LD standard and smart data models[10] for better interoperability. As part of the IDS standard, the IDS information model[5] was employed for the metadata descriptions, tailored to the NGSI-LD specification to enable data access with the highest possible granularity.

In this regard, it has to be noted that the IDS information model hierarchically organizes the metadata descriptions in IDS-catalogs that comprises IDS-resources with multiple IDS-representations, which are further composed of IDS-artifacts, typically one per file or data access endpoint. Our approach considers one single IDS-catalog, where each context information type from an NGSI-LD CB is mapped into an IDS-resource, creating multiple IDS-artifacts for each of the entities of that type and their historical record. Additionally, an IDS-contract and a set of IDS-rules apply to each of the IDS-resources. This way, access and usage control rules can be mapped to each NGSI-LD type.

Figure 2 presents an example of the mapping for a data provider offering information about a fleet of vehicles (i.e., employing the NGSI-LD type FleetVehicleStatus). There is one IDS-artifact for accessing the list of monitored vehicles (light green); multiple IDS-artifacts, depending on the number of vehicles, for getting the latest context information (light yellow) and their historical record (i.e., time-series of values) for each of them (light orange). Furthermore, additional uniform resource locator query strings are forwarded to the NGSI-LD CB in the back end, enabling further functionality (e.g., querying for a specific period when retrieving historical data).

CO_2 FOOTPRINT MONITOR

To validate the proposed architecture, we have carried out a PoC pilot to interconnect data spaces from Europe and Japan. Due to its relevance to climate change, we have implemented a CO_2 footprint monitor. This application consumes different kinds of mobility-related data to estimate vehicles' CO_2 emissions based on the ODALA Mobility Data Toolkit (MDT),[11] a tool to track emissions based on vehicle metrics, using estimations for different transport means provided by European Environment Agency.

On the European side, an IDS-based data space was established within the framework of SmartSantander[12] and populated with heterogeneous open and private

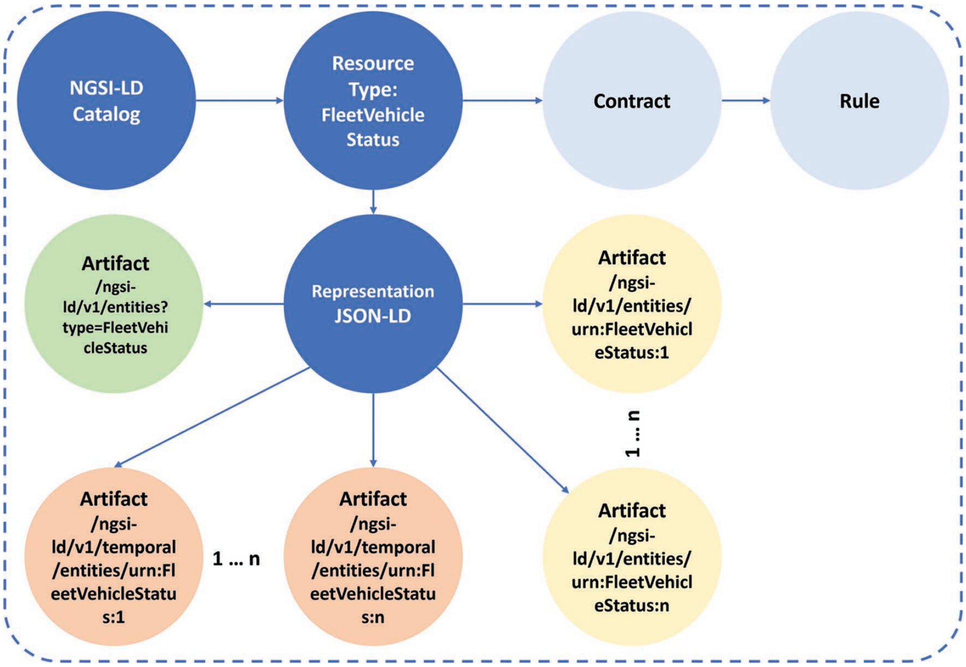

FIGURE 2. Mapping of IDS information model on NGSI-LD for FleetVehicleStatus entity types.

datasets, including real-world information from a local bus company about buses' position, routes and speed, updated in real-time. The connector selected was the IDS-certified Data Space Connector (DSC). On the Japanese side, a CADDE-based data space was used, hosting information on vehicles' position, engine type, and vehicle consumption metrics, framed within the Strategic Innovation Program (SIP).[4] The whole implementation architecture is depicted in Figure 3.

The CO_2 footprint monitor includes the MDT core library, a Scorpio Broker,[13] and a dashboard. It is fed with information retrieved from different data providers at both CADDE and IDS data spaces, being imported into the Scorpio Broker in NGSI-LD, following the vehicle and FleetVehicleStatus smart data models.[10] All these mobility-related data are being processed continuously, estimating the CO_2 emissions per vehicle and passenger, enriching the information obtained from both data spaces. Everything is shown on a dashboard as a multigraph visualization of vehicles CO_2 emissions. As a PoC pilot, the CO_2 footprint monitor has demonstrated the potential to serve as a tool for city managers, as well as for other use cases (e.g., the monitoring of CO_2 emissions in manufacturing across a worldwide supply chain) thanks to its ability for cross-data space interworking.

DISCUSSION AND LESSONS LEARNED

The pilot has demonstrated the feasibility of an interworking solution to connect data spaces supported by the dominant technologies in different world regions. However, there are still certain limitations that need to be addressed to fully realize the potential of decentralized self-sovereign data spaces.

Data space standards require that all participants are fully compliant with their respective guidelines (e.g., identity management, information models, data exchange protocols, and so on). Consequently, using different technology stacks is technically feasible, but it creates a disruption in the trust domain established among providers and consumers, as there is no real end-to-end communication under the umbrella of a single data space. Therefore, international cooperation is essential to enable fully compliant data space interworking, agreeing on common guidelines. Data space initiatives should converge to common identity management mechanisms and procedures to facilitate it.

While initiatives such as the Data Space Business Alliance (DSBA) aim to tackle this limitation, they are still at the proposal stage phase and may not cover all standards in the future, especially those from different regions. In this regard, new generations of data spaces emphasize the use of decentralized SSI, which allows users to avoid relying on central identity providers governed by each data space. By relying on common trust anchors, identity management could be seamlessly integrated across multiple data spaces, allowing for greater interoperability and transversal data exchange. Similarly, extendable common information models for resource metadata description shall be agreed upon in the future, beyond the limits of regional data space frameworks, easing the discovery process.

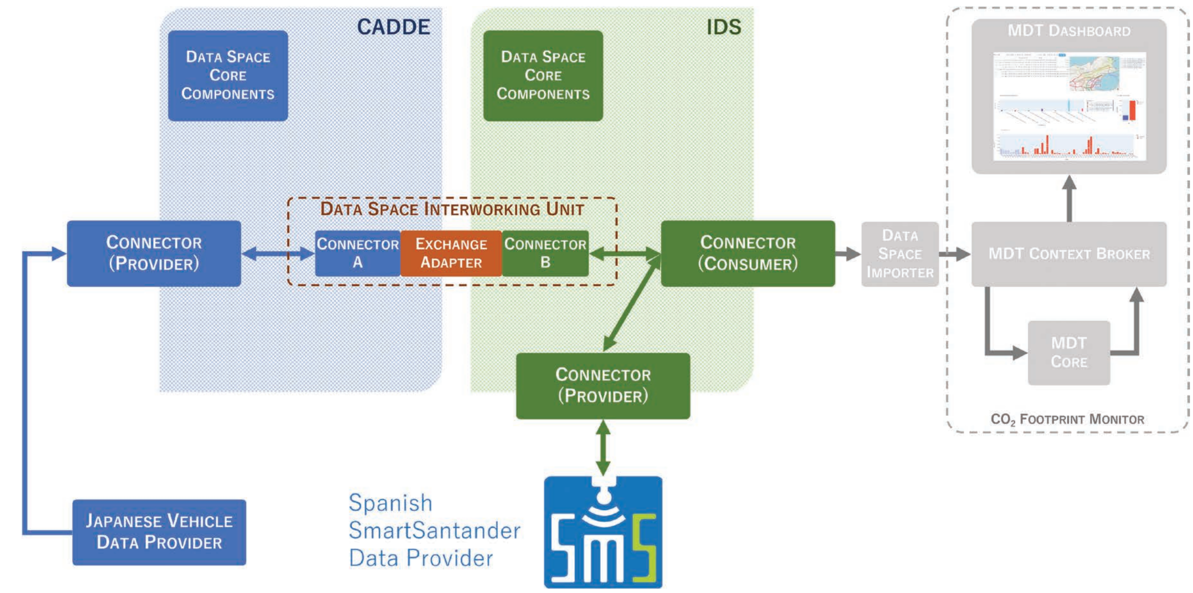

FIGURE 3. Holistic view of the CO_2 footprint monitor.

Furthermore, none of the reviewed data space technologies impose any restrictions regarding back-end APIs and information models for the data exchanged. Nonetheless, the advantages of using a common model are clear for better interoperability. During the pilot, it was decided to utilize NGSI-LD and the smart data models, to ensure interoperability between participants. However, current IDS-certified implementations are constrained by the IDS information model, as it only allows access to information sources upon request, as they are geared toward dataset and file exchange, excluding asynchronous subscription-based data access, which is particularly critical for real-time Internet of Things applications.

CONCLUSION

Data spaces are attracting a lot of attention as the reference ecosystems to enable trustworthy decentralized data exchange among stakeholders. However, developed standards are being established in isolated regions (e.g., Europe, Japan, United States, and so on), thus limiting the scope of their data spaces to their regions. We have highlighted the importance of enabling the interworking across data spaces based on different technological stacks to promote a truly global data economy.

The technical solution presented in this article facilitates data sharing across data spaces employing different technological stacks. Moreover, a case study to monitor CO_2 emissions in a global scenario has been developed, employing data that was available at two intercontinental data spaces, based on IDS and CADDE. The PoC validation demonstrated the feasibility of such interworking, but also that there are still key limitations that have to be tackled. These are mainly related to the need for cooperation in terms of building trust between providers and consumers in different regions.

ACKNOWLEDGMENTS

This work was supported in part by the Cross-Ministerial Strategic Innovation Promotion Program (SIP) Phase 2/Cyber Space Infrastructure Technology Utilizing Big Data and AI led by the Council for Science, Technology, and Innovation (CSTI) of the Japanese Government. It was also supported in part by the project SEDIMARK (SEcure Decentralized Intelligent Data MARKetplace) from the Horizon Europe Program (Grant 101070074) and by the Spanish State Research Agency (AEI) through the project THROTTLE (TrustwortHy uRban mObility daTa markeTpLacE, Grant TED2021-131988B-I00).

REFERENCES

1. G. Micheletti, N. Raczko, C. Moise, and D. Osimo, “Data market study 2021–2023 - First report on policy conclusions,” Jun. 9, 2022. Accessed: Jan. 17, 2025. [Online]. Available: https://ec.europa.eu/newsroom/dae/redirection/document/87523
2. “International data spaces association: Home.” International Data Spaces. Accessed: Jan. 17, 2025. [Online]. Available: https://internationaldataspaces.org/
3. “Home - Gaia-X: A federated secure data infrastructure.” Gaia-X. Accessed: Jan. 17, 2025. [Online]. Available: https://gaia-x.eu/
4. N. Koshizuka and H. Mano, “DATA-EX: Infrastructure for cross-domain data exchange based on federated architecture,” in *Proc. IEEE Int. Conf. Big Data*, Dec. 2022, pp. 6145–6152, doi: 10.1109/BigData55660.2022.10020855.
5. C. Mader, J. Pullmann, N. Petersen, S. Lohmann, and C. Lange-Bever. “International data spaces information model.” GitHub Pages. Accessed: Jan. 17, 2025. [Online]. Available: https://w3id.org/idsa/core
6. “Dataspace protocol 2024-1.” IDS Knowledge Base. Accessed: Jan. 17, 2025. [Online]. Available: https://docs.internationaldataspaces.org/ids-knowledgebase/v/dataspace-protocol/
7. “Gaia-X Architecture Document - main version (0caf1aee).” Accessed: Jan. 17, 2025. [Online]. Available: https://gaia-x.gitlab.io/technical-committee/architecture-working-group/architecture-document/
8. F. Cirillo, G. Solmaz, E. L. Berz, M. Bauer, B. Cheng, and E. Kovacs, “A standard-based open source IoT platform: FIWARE,” *IEEE Internet Things Mag.*, vol. 2, no. 3, pp. 12–18, Sep. 2019, doi: 10.1109/IOTM.0001.1800022.
9. “Industry specification group (ISG) cross cutting context information management (CIM).” ETSI. Accessed: Jan. 17, 2025. [Online]. Available: https://www.etsi.org/committee/cim
10. FIWARE. “Smart data models – A global program led by.” Smart Data Models. Accessed: Jan. 17, 2025. [Online]. Available: https://smartdatamodels.org/
11. NEC Laboratories Europe. “ODALA mobility data toolkit.” GitHub. Accessed: Jan. 17, 2025. [Online]. Available: https://github.com/nec-research/mobility-toolkit
12. L. Sanchez et al., “SmartSantander: IoT experimentation over a smart city testbed,” *Comput. Netw.*, vol. 61, pp. 217–238, Mar. 2014, doi: 10.1016/j.bjp.2013.12.020.
13. NEC Laboratories Europe. “Scorpio NGSI-LD broker.” GitHub. Accessed: Jan. 17, 2025. [Online]. Available: https://github.com/ScorpioBroker/ScorpioBroker

JUAN RAMÓN SANTANA is an assistant professor at the Universidad de Cantabria, 39005, Santander, Spain. His research interests include data management platforms, Internet of Things, and wireless sensor networks. Santana received his Ph.D. degree from the Universidad de Cantabria. Contact him at jrsantana@tlmat.unican.es.

LUIS SÁNCHEZ is an associate professor at the Universidad de Cantabria, 39005, Santander, Spain. His research interests include Internet of Things-enabled smart cities, meshed networking, and data management platforms. Sánchez received his Ph.D. degree from the Universidad de Cantabria. Contact him at lsanchez@tlmat.unican.es.

MARTIN BAUER is a principal standardization engineer at NEC Laboratories Europe, 69115, Heidelberg, Germany. His research interests include data management platforms for real-world information and data spaces and FIWARE, NGSI-LD standardization in ETSI ISG CIM and Gaia-X. Bauer received his Ph.D. degree from the University of Stuttgart. Contact him at martin.bauer@neclab.eu.

BENJAMIN HEBGEN is a senior software engineer at NEC Laboratories Europe, 69115, Heidelberg, Germany. His research interests include Internet of Things-related research activities such as context-aware advertising and digital twins in the smart cities context. Hebgen received his B.Sc. degree in computer science from Technische Hochschule Mannheim. Contact him at benjamin.hebgen@neclab.eu.

ERNÖ KOVACS is the senior manager for the Data Ecosystem and Standards Group, NEC Laboratories Europe, 69115, Heidelberg, Germany. His research interests include context brokering, cloud–edge computing, and smart cities. Kovacs received his Ph.D. degree from the University of Stuttgart. Contact him at ernoe.kovacs@neclab.eu.

SATSUKI HAMAGUCHI is an assistant manager at NEC Corporation, Tokyo, 108-8001, Japan. Her research interests include technology strategy of data integration, being active in new business development for data spaces, and smart cities. Hamaguchi received her M.Sc. degree from the Department of Mechanical Engineering at Nihon University. Contact her at wild-elf-1@nec.com.

YURIKO NOMURA is a senior manager at NEC Corporation, Tokyo, 108-8001, Japan. Her research interests include new business promotion and technology strategy for data-driven smart cities. Nomura received her B.Sc. degree from the Department of Electrical Engineering at Keio University. Contact her at y-nomura@nec.com.