Validating Architecture for Real-world Deployment of Data Space Technology and Outlook for Practical Use

3.1 Introduction

As our digital society evolves, the conventional centralized platform model, under which vast amounts of data are collected in one place for management and use, is increasingly being complemented by a new approach that is drawing attention: the "data space." A data space enables autonomous, decentralized interconnection while preserving data sovereignty. This growing interest is driven by a fundamental need: although the ability to use data across organizational boundaries is now crucial, organizations also need to eliminate by design such risks as the leakage of trade secrets and the use of data for unintended purposes, thereby reconciling trust and control in the way data is shared.

These developments originated in Europe, where frameworks for trust-based data collaboration have been put forward through pioneering initiatives such as Gaia-X and International Data Spaces (IDS). In Japan as well, the drive toward real-world adoption is gaining speed, as efforts advance to establish the reference architecture known as ODS-RAM through collaboration among industry, government, and academia.

To verify the technical effectiveness of this new framework for data collaboration, we conducted a proof of concept (PoC) in April 2026. Our primary objective was to establish foundational technologies for enabling secure data collaboration in distributed environments while preserving data sovereignty. Specifically, we built an interconnection environment that combined an ODS-compliant data transfer module with the IIJ Cloud Data Platform and validated its effectiveness by integrating authorization mechanisms such as Relationship-Based Access Control (ReBAC). Building on the results of this work, we also examined the technical feasibility of extending this approach to domains requiring more stringent governance, with a view to future integration with advanced foundational technologies for data protection currently being developed as part of the K Program (NEDO)[*1].

This article presents the basic architecture of the data space technology we examined in the PoC, its specific implementation, and the PoC scenarios we designed with a view to enhancing the cyber resilience of telecommunications carriers, together with the findings obtained.

3.2 Architecture for Realizing Data Spaces

What defines a data space in technical terms is not the centralized accumulation of data, but rather collaboration among participants based on autonomous trust. This section outlines the core architecture of a data space and explains the role of each component.

3.2.1 Trust-based Collaboration and the Preservation of Data Sovereignty

The biggest feature of a data space is that it provides data sovereignty, under which data providers retain control over their data. In conventional data-sharing models, once you entrusted data to a platform, you had to delegate control over its use to that platform. In contrast, a data space introduces policy-based access control to establish a mechanism for usage control, whereby providers define the terms of use and recipients are required to comply with them through technical enforcement.

This is supported by a trust framework that extends across organizational boundaries. ODS-RAM places particular emphasis on the principle of Trust by Design (trust is not assumed but established through system design). Under this approach, even participants that do not have a fixed pre-existing relationship can initiate secure data exchange by proving their identities through verifiable credentials. Note, however, that an authenticated entity is not necessarily a trusted one. What matters is that the parties agree on the level of trust required for the relevant risks and intended use, and then combine measures such as verification, digital signatures, and trusted timestamps to build a trust structure that is fit for purpose.

3.2.2 Four-layer Architecture of a Data Space

In our PoC, we implemented a minimal configuration, referring to the layered structure defined by ODS-RAM, Japan's standard reference model. Conceptually, the architecture of a data space consists of the four layers described below. Because the layers are mutually independent and loosely coupled, the architecture allows the necessary layers to be

*1 Key and Advanced Technology R&D through Cross Community Collaboration Program (NEDO), Data no hogo to ryutsu no jidoka gijutsu no kenkyu kaihatsu [Research and development of automation technology for data protection and distribution] (https://www.iij.ad.jp/news/pressrelease/2023/1108.html, in Japanese).

implemented selectively according to the maturity of the domain and use case.

■ **L1: Data Layer**

This layer addresses usage control, protection against data tampering, and data quality. A key requirement is that the actual data remain within the provider's own infrastructure, allowing the provider to exercise usage control at its own discretion.

■ **L2: Transaction Layer**

This layer governs the entire transaction process between the provider and the consumer in a manner that is independent of data format and transfer method. In addition to physical transfer, it handles routing, automated policy enforcement, and the technical enforcement of agreed terms. These functions are performed by connectors deployed at the edge of each participant's environment in a distributed design that is fundamentally different from a centralized hub-based architecture.

■ **L3: Identity Layer**

This layer handles participant authentication and authorization. It provides credentials in a verifiable form and applies access control based on the terms of use set by the data provider. ODS-RAM breaks identity down into three elements: verification of real-world existence, authentication, and authorization. And it treats successful technical authentication and the decision to trust an entity as separate issues.

■ **L4: Semantics Layer**

This layer deals with addressability and semantics. By making metadata accessible, it enables interoperability through discovery, interpretation, and use of data across organizations.

3.2.3 How Autonomous, Distributed Collaboration Works

When the layers described above operate together, they enable high-trust data collaboration through the following process, without relying on a centralized platform.

■ **Dynamic Binding of Identity and Policy**

At L2 (the transaction layer), the usage policies defined by the data provider in L1 (the data layer) are dynamically matched against the participant attributes proven through L3 (the identity layer). As a result, even in the absence of a prior agreement with a given counterparty, access can be authorized securely based solely on real-time attribute verification.

■ **Enforcing Policy at the Edge**

A core feature of a data space is that the connector located at the point of data egress (the edge) performs the L2 role and physically enforces policy at that point. By making authorization decisions and applying controls before data leaves the provider's administrative domain, it prevents the provider from losing control over the data and ensures data sovereignty by technically enforcing terms of use even after the data has been handed over.

When these mechanisms work together, participants can manage their data autonomously while forming a "network of trust" in which they share information only with the parties they need to, only when they need to.

3.3 Design and Implementation of the PoC System

To implement the architecture described in Section 3.2 for our PoC, we built a demonstration environment by interconnecting an external ODS-RAM-compliant data space infrastructure service with our own service assets. For this phase, we adopted a minimal configuration aimed at rapidly assessing technical viability, so of the architecture's four layers, the implementation focused on L1 (the data layer), L2 (the transaction layer), and L3 (the identity layer). L4 (the semantics layer), which is responsible for defining data semantics, was therefore excluded from the scope of this PoC. Cross-cutting design elements such as governance, security, and trust (Trust by Design) were considered only at the level of basic design guidance. Below we describe the implementation structure in detail, including how each layer's functions were mapped to specific components.

3.3.1 Trust Model and Authorization Control in the Identity Layer (L3)

In implementing the identity layer (L3), we used an external infrastructure service as a trust anchor for managing the root of trust. For the authentication process in this PoC, we adopted OpenID Connect (OIDC) using Keycloak and built a mechanism for verifying the legitimacy of participating entities. For authorization control, we implemented Relationship-Based Access Control (ReBAC) across L2 and L3, enabling dynamic authorization checks based on inter-organizational relationships and privileges. By linking authorization information with an external infrastructure service compliant with industry-standard specifications, we verified that this configuration can support future interoperability across multiple domains.

3.3.2 Implementation of the Data Collaboration Platform and L2 Connector

For the connector function at the core of the transaction layer (L2), we adopted a data transfer module developed in accordance with the ODS L2 specification. Complementing this, our cloud-based data collaboration platform (the IIJ Cloud Data Platform) acts as an intermediary between the L1 data source and the L2 transfer module, providing advanced hub functionality to facilitate seamless data collaboration. Their respective roles are as follows.

■ Policy Enforcement

The L2 data transfer module evaluates, in real time, authorization information from the L3 infrastructure against the provider's usage policies. By ensuring that data is sent only to authorized recipients and only under the specified conditions, it serves as a technical barrier for maintaining the provider's data sovereignty.

■ Interface Abstraction and Push-Notification Handling

The IIJ Cloud Data Platform abstracts away interface differences across individual L1 data sources and gives the L2 module a unified HTTP API for data access. And to complement the inherently pull-based communication model of the data space, it also handles push notifications that instruct the counterparty to retrieve data, thereby enabling a practical data collaboration process.

3.3.3 Integration of Backend Systems and the Data Layer (L1)

For the data layer (L1), anticipating future integration with operational management systems, we used a dataset consisting of a subset of equipment information. A key design point in our implementation is that it dynamically extracts and provides only the data required through the data collaboration platform described above, without directly modifying the existing business systems. With this mechanism, we confirmed that our approach can quickly and securely make static data stored in existing systems available and usable as dynamic assets within the data space, while maintaining confidentiality.

3.3.4 Access Control and Future Extensibility

At the boundary with the data space, we implemented an access control mechanism that combines the L2 transfer module with the L3 authorization mechanism (ReBAC). This mechanism allows only requests authorized by the L3 infrastructure to pass, providing robust protection against unintended disclosure to third parties. Based on the fit-for-purpose trust structure described in Section 3.2.1, it ensures data trustworthiness by combining encryption of communication channels with strict verification of authorization information.

The L3 authorization mechanism we adopted is general-purpose enough that it can also be referenced by components other than the L2 modules. As such, in the future we expect it to enable flexible integration with IIJ's other services, as well as with the advanced foundational technologies for data protection being developed as part of Japan's K Program (NEDO). By further advancing the outcomes of this research and integrating technologies that more effectively achieve both data confidentiality and usage with policy-control and access-control technologies that govern control across organizational boundaries, it may be possible to develop this into a data distribution platform that combines security and flexibility even in domains where stricter governance is required.

3.4 Verifying Technical Effectiveness Through PoC Scenarios

To verify the value that data space technology can deliver, we conducted a PoC assuming a wide-area telecommunications carrier equipment data space.

3.4.1 The Need to Strengthen Cyber Resilience Through Wide-Area Equipment Collaboration

As attacks on telecommunications infrastructure grow more sophisticated, it is becoming difficult for any single operator alone to take steps to manage risks across the entire supply chain. In particular, when it comes to understanding vulnerabilities in equipment spanning organizational boundaries and the sharing of maintenance status, information silos and delays in information sharing pose structural challenges. In this PoC, we aimed to solve these challenges through autonomous information linkage.

3.4.2 Improving Maintenance Processes via Attribute-Based Access Control

In the first scenario, we focused on identifying equipment subject to maintenance updates between a telecommunications carrier and its partners.

Ensuring that End of Service Life (EOSL) is not overlooked is essential not only for avoiding risk, but also as a basis for making timely decisions on strategic system refreshes that ensure business continuity and adaptation to the latest technology.

In the PoC, we confirmed a process for selectively disclosing and obtaining data within the scope authorized based on attributes (credentials). This confirmed the practical utility of accelerating the identification of equipment requiring updates and enabling planned capital investment.

3.4.3 ynamically Sharing Support Information Using a Chain of Trust

In the second scenario, we tested a process for reliably delivering vendor-issued vulnerability information to operations personnel.

Here, we built a flow in which those who need the information obtain it at the time they need it via the data collaboration platform, with information authenticity being guaranteed. We confirmed that this eliminates the degradation and delays that come with passing information through human hands, minimizing lead time in the initial response to an incident.

3.4.4 Summary of PoC Results

Through this series of scenarios, we confirmed that it is indeed possible to implement systems that provide both confidentiality and transparency (controlled transparency), such that, based on attributes, data is disclosed only to those meant to see it, while preserving data sovereignty. The main results of our PoC are as follows.

■ Autonomous Trust Relationships Shorten Collaboration Lead Times

Sharing equipment data across organizations has traditionally carried heavy coordination costs, such as those involved in signing individual NDAs and building dedicated networks. In this PoC, we combined attribute-based authentication on the L3 platform with policy enforcement by L2 connectors to create a flow that pulls in data from trusted parties instantly and securely. This provides a foundation for rapid decision-making across organizational boundaries in emergencies such as cyberattacks.

■ Making Data an Asset While Leveraging Existing Systems

We showed that the existing production asset, the Unified Operations Management service (UOM), can be adapted for use in a data space solely through on-demand extraction via a connector, without major modifications. For infrastructure operators with vast legacy assets, this is a realistic and highly effective way to move toward the latest in distributed data collaboration while protecting existing investments.

■ Establishing a Starting Point for Improving Resilience Across the Entire Infrastructure

Configuration and vulnerability information that had been siloed within individual operators can be loosely linked through the data space to the extent needed, opening a path to visualizing risk across the entire supply chain in near real time. This presents a new form of data collaboration, one that connects together the measures taken by individual operators and raises the cyber resilience of social infrastructure as a whole.

3.5 Technical Considerations and Future Outlook

While our PoC confirmed the effectiveness of data spaces, it also revealed the challenges and prospects involved in implementing them as wide-area social infrastructure and building a sustainable ecosystem around them.

3.5.1 Technical Challenges for Real-World Deployment and Advancing the Trust Model

To improve practical utility further, we first need to solve technical challenges around the dynamics of data distribution and sharing of semantics.

The first challenge is implementing an event-driven delivery mechanism that improves the timeliness of information. In addition to the current on-demand data extraction, we need a mechanism whereby specific occurrences (events) trigger connectors to autonomously push information out to the relevant organizations.

This would be useful not only for incident response tasks, such as detecting equipment anomalies or discovering vulnerabilities, but in any use case in which information needs to be shared without delay, such as changes in inventory levels or environmental sensors exceeding thresholds. It would broaden the scope of what data spaces can do, from static collaboration, where you actively go and fetch information, to dynamic collaboration, whereby the information you need arrives at just the right time.

The second challenge is standardizing semantics to automate the interpretation of data across organizations, which was outside the scope of this PoC. As a data space grows,

differences between the vocabularies independently defined by each participant and cases in which different organizations assign different IDs to the same entity end up becoming structural obstacles to collaboration. Effective approaches here are dynamic mapping based on a common ontology for the former, and the assigning of global identifiers using DIDs (Decentralized Identifiers) for the latter. What matters in both cases is that participants can achieve interoperability without changing their own existing definitions or IDs. Combined with the event-driven delivery described above, this becomes a foundation that supports both the distribution of data and its interpretation and identification as a unified whole.

To undergird these information dynamics and shared semantics, we must progressively advance the trust model and build the institutional frameworks that support it. In our PoC, we adopted a community-based model that uses a specific authentication infrastructure as its trust anchor. The next technical milestone is to build on this and extend it into a more autonomous, decentralized trust model based on VCs (verifiable credentials).

A key part of this process will be mechanisms that go beyond simply verifying technical signatures to dynamically determine who has vouched for a given credential and against what criteria. Specifically, we will need a function like the Clearing House proposed in Gaia-X, which verifies participant conformance and records and audits transaction evidence, together with a federated catalog function that catalogs data assets and policies and shares them among participants.

In such mechanisms, defining governance as in who serves as the trust anchor (the root of trust) goes hand in hand with technical implementation. By creating an environment in which both providers and consumers of data can assess and reference dynamically changing trust levels under a common framework, we will be able to maintain trustworthiness even as the number of participating organizations grows, while dramatically reducing interconnection costs.

3.5.2 Building a Data Ecosystem That Creates Value

If data spaces are to become a sustainable framework rather than ending as one-off demonstrations, they must evolve from mere venues for exchange into ecosystems that turn the information flowing through them directly into real business value. To make this happen, it will be crucial to design a structure that facilitates independent participation by not only those who provide and obtain data but also third parties who create value on top of it (application developers, analytics service providers, and the like).

As European precedents show, one effective step is to build marketplaces that offer not just the data itself but also analytics capabilities (services) as a package for processing and analyzing the data. The trouble with data provision and use alone is that participants struggle to see any direct benefit, so it is hard to generate motivation that outweighs the cost of connecting. Catena-X tackles this problem by limiting the applications and services listed on its marketplace to certified offerings only, giving participants an environment they can use with confidence and without the burden of vetting vendors. This lowers the barrier to connecting, and it also drives network externalities, such that as the number of participants increases, the data become more diverse and the analytics become more accurate.

Standardized on-ramps like KITs (Knowledge, Implementation, Tooling) are also an important way to encourage third parties to enter the ecosystem. By packaging the technical specifications, implementation guidelines, and tools needed for each use case, they let application developers join the ecosystem in a standards-compliant way without having to start from scratch with their own research and individual negotiations.

And to support the healthy growth of this ecosystem, guaranteeing the neutrality of the underlying technology stack is also crucial. Building on open source software (OSS) components that do not depend on any particular vendor, and on open standard specifications that anyone can conform to, ensures fair connection opportunities for participating organizations. Ensuring this neutrality and transparency encourages diverse stakeholders to join with confidence, broadens the pool of those creating value on the platform, and drives the ecosystem's self-reinforcing growth. Offering small and midsize organizations a low-cost option for connecting via managed services is another essential element for broadening the ecosystem base.

3.5.3 Prospects as General-Purpose Infrastructure and Convergence with AI

Just as the Internet once removed the barriers to communication and transformed the structure of society, data spaces have the potential to drive the evolution of data usage from

a special undertaking into a standard operational foundation. For this to happen, we will need to see progress on technical standardization as well as widespread awareness of the real benefits that participants gain by connecting to a data space.

As a step toward real-world deployment, a realistic approach is to start with closed, community-based trust within a specific industry and among parties with shared objectives, building trust relationships grounded in operational rules. From there, gradually shifting and expanding to general-purpose trust based on machine verification using verifiable credentials (VCs) should ensure interconnectivity between different ecosystems, such that data spaces grow into a truly open data distribution platform.

This evolution of the trust infrastructure will further accelerate deep convergence with AI technology. With large language models (LLMs), for example, we can now envision the prospect of building environments in which a question posed in natural language safely and automatically retrieves and integrates the necessary information, without the need to think about complex policy settings or data mappings. And if general-purpose trust provides machine-verifiable guarantees of data provenance and rights, the trustworthiness of AI-generated output will also rise dramatically.

Meanwhile, data spaces will also be a vital foundation for AI, because an environment in which high-quality data of clear authenticity flows across organizations will have a direct impact on improving the training accuracy of and ensuring governance over AI models. Once data flowing across organizational boundaries becomes the norm, this will surely create fertile ground for the creation of new value, not just in terms of strengthening cyber resilience but in all sorts of areas, including carbon neutrality and supply chain optimization.

3.6 Conclusion

Our PoC was an attempt to put the data space concept into practice in telecommunications infrastructure, a domain that demands high reliability, as a step toward realizing next-generation data collaboration. The insights we have gained go beyond the challenges of any specific industry and point toward the sort of flexible form that data distribution should take going forward.

Our aim is to bring about a society in which data is used safely and efficiently in a way that is trusted, unfettered by organizational boundaries. To achieve this, we will deploy advanced technologies such as AI to simplify implementation, opening the door to a future in which the new infrastructure of the data space sees widespread adoption and ongoing development as an open service freely available to everyone.

At IIJ, we will continue to work with stakeholders on efforts to bring about a new data-driven society, in terms of both technology and services.

Yasushi Toriumi

Business Development Office, Enterprise Sales Unit, IIJ
Since joining IIJ, Mr. Toriumi has worked primarily as a project manager in the field of system integration, launching solution businesses aimed at expanding the integration business. He currently serves as a technology strategist in the Business Development Office, pursuing initiatives to bring about a data-driven society, mainly through the development and utilization of data distribution platforms.