The power of data spaces

A roadmap for digital transformation in the European public sector

Drawing on the extensive experience of NTT DATA and Amazon Web Services (AWS), this perspective explores the role of data spaces in driving the digital transformation of public administrations across Europe.

The aim is to equip policymakers, public-sector managers and other interested audiences with the knowledge and tools to realize the full potential of data spaces in enhancing digital public services and driving innovation in the data-driven landscape.

We outline the context and benefits of data spaces in the public sector and share real-world examples, including the Public Procurement Data Space (PPDS) and the European Mobility Data Space (EMDS), to demonstrate the transformative impact of data spaces across diverse domains.

We then delve into the key considerations and challenges that public-sector organizations must navigate when adopting this new paradigm, covering areas such as data quality, governance, digital service adoption, and ensuring data sovereignty and trust.

Building on this foundation, a detailed overview of high-level data-space architecture and core components is provided. This includes an examination of the diverse stakeholder roles, trust frameworks, data-exchange protocols and connector technologies that enable effective data exchange.

We conclude with practical guidance on hosting, securing and architecting data spaces using AWS services and infrastructure, and share 10 key lessons learned from developing and implementing impactful data spaces.

aws PARTNER Competency

Contents

The data-space landscape in the European Union (EU)

Since data is vital to the global economy, the European Union (EU) requires resilient and competitive data-management capacities to serve both the private and public sector.

A data space is a distributed system that is defined by a governance framework to allow secure, trustworthy data transactions between participants while supporting data sovereignty.

With secure, standardized digital infrastructure, shared rules and policies, and data-governance mechanisms, data spaces make data management more resilient by keeping data secure, accessible and compliant with regulations. These trusted "data exchanges" allow stakeholders to access data and data-based services securely and transparently.

In the EU, the Data Spaces Support Centre (DSSC) coordinates and supports the development of data spaces in Europe. Important stakeholders in this field include, the International Data Spaces Association (IDSA), the FIWARE Foundation, Gaia-X and the Big Data Value Association (BDVA). Together, they form the Data Spaces Business Alliance (DSBA).

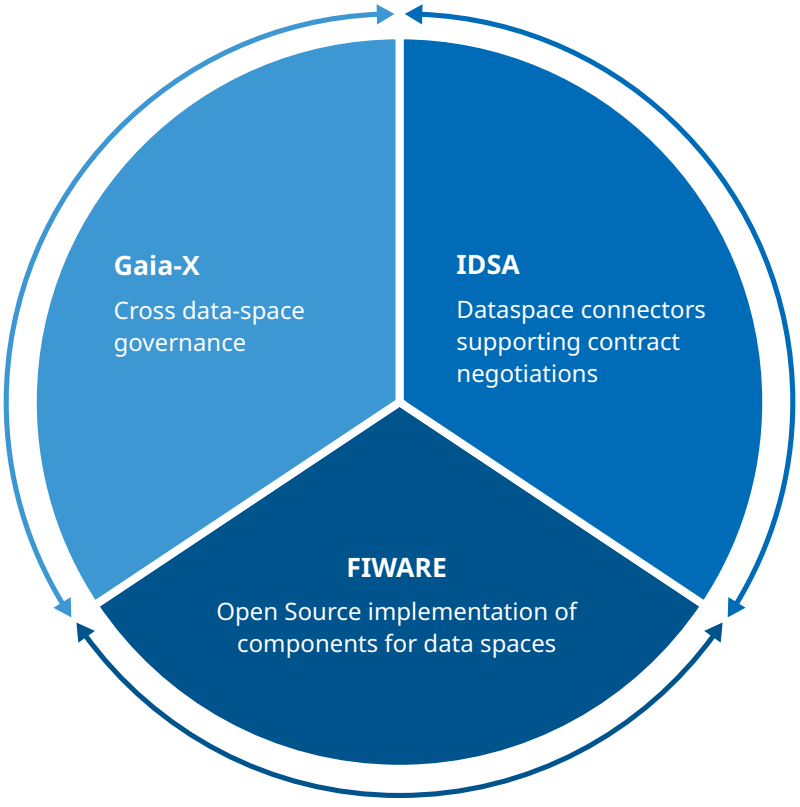
To support data access and interoperability among European data spaces, the European Commission has launched the Simpl initiative. Simpl is an open-source, smart and secure middleware platform that will enable the creation of EU data spaces and facilitate interoperability between existing data spaces.

Simpl will first be deployed in key areas of the public sector such as healthcare, public procurement, linguistic and cultural heritage, mobility and earth observation.

The main vision of the Simpl initiative is to facilitate data sharing across data spaces and scale data from the edge to the cloud across interconnected domains.

BDVA

Providing advice and general consultation, and endorsing results



Source: DSBA Technology Convergence

Data spaces for the public sector

Data spaces are emerging as a powerful tool for enhancing data sharing, integration and collaboration across diverse sectors and organizations. Although their development is still in the early stages, with open questions about scalability, boundaries and trust-enforcement mechanisms, they are gaining attention from industry and business.

The existing focus has been on the industrial sector, where data sharing remains a significant challenge because of the prevalence of private information. However, public administration data spaces hold immense promise, as they can leverage the wealth of available public data to enhance transparency, strategic decision-making and citizen engagement.

> “ In this document, the terms “public sector” and “public administrations” refer to government entities, agencies and institutions at national, regional and local levels that are responsible for providing public services and managing of public affairs.

In the public-sector context, data spaces are particularly relevant and important, as they offer a solution to the persistent challenge of fragmented information systems and data silos. By harnessing the potential of data spaces, public-sector entities can play a critical role in improving the quality and efficiency of public-service delivery.

Data spaces enable the seamless transfer of vital information between digital services, leading to enhanced automation, proactive service provision and better-informed policymaking. They also foster opportunities for the creation of new value-added services by facilitating the controlled exchange of data between the public and private sectors.

3 reasons data spaces are important in the public sector

1. Data spaces enable the seamless transfer of information between digital services through a trust framework.

In practice, this means data can be exchanged between different systems and organizations, ensuring the secure control of data access and usage.

This translates into reduced times for administrative procedures and minimizes the transfer of data outside its source.

2. Data spaces align with the ongoing digital transformation of the public sector, particularly in managing base registries in, or close to, real time.

They facilitate the generation of value-added services by fostering a massive exchange of data while sharing only precise and necessary information for administrative processes.

This trajectory reflects a strategic shift toward future systems of records and digital services in public administrations. These systems will operate based on events that, when properly orchestrated, will yield opportunities for further automatization, knowledge generation and the proactive delivery of digital public services.

Ultimately, this transformation aims to center public administrations on the citizen and society while fostering greater engagement and efficiency. A practical example is the seamless exchange of health-record data between actors while complying with usage policies and agreements independently of their underlying technological stacks.

3. Data spaces address the challenge of fragmented information systems and data in the public sector.

Previously, this fragmentation hindered interactions among citizens, organizations and public-services organizations, needing a significant redesign and rebuilding of the public-service IT ecosystem.

Data spaces offer a solution by efficiently connecting existing data and information silos in a meaningful way, leveraging the existing ecosystem and avoiding the need for a complete redesign. This makes data spaces a pivotal tool for streamlining interactions and enhancing the effectiveness of public services.

Ultimately, the implementation of data spaces in the public sector places citizens at the forefront, streamlining their interactions with administrative services while safeguarding personal data sovereignty.

Potential benefits of adopting data spaces in the public sector

The main advantages of data spaces in the public sector are:

1. **A paradigm shift to data-driven policy development:** Access to a range of data sources can allow policymakers to formulate evidence-based policies that are effective and targeted.
2. **Improved decision-making in public administrations:** Access to and sharing of comprehensive information enable policymakers to make informed decisions.
3. **Efficient service delivery:** The integration of master data, registries, existing or legacy digital public services, and different datasets (data catalogs) facilitates the creation of new public services. For example, in a city, sharing necessary information can lead to the establishment of a complete and autonomous civil registration service, including all related procedures.
4. **Data-economy enabling:** Public administrations have a vast amount of information that is generated or collected by digital public services. Allowing private operators to consume and reuse such information can lead to the creation of value-added services and new business models.
5. **Enhanced transparency:** Making (open) data available through the EU and public administrations, combined with the publication and updating of data, will make public services and public administration operations more transparent and accountable. Ultimately, increased transparency creates trust in public management.
6. **Cost savings:** Data-driven decision-making can lead to efficient resource allocation, reducing costs and improving the overall financial management of public institutions.
7. **Citizen engagement:** More citizens will participate in exchanging their data with public administrations that provide and consume data. Citizens will be able to decide which data they share or keep private under certain conditions.
8. **Compliance with data regulations:** A critical aspect of public administrations is the development of digital public services that are compliant with regulations, especially those that relate to data protection.
9. **Research and innovation:** Giving researchers access to large, open datasets will help them to advance scientific knowledge.

The main idea behind data spaces is distributing and federating data management across countries and sectors (see Figure 1). For this to work, and in the case of the public sector, the data economy must be based on increasing the value of the data shared by incorporating newer services on top of the raw data (for example, by enriching it with metadata, applying machine learning and applying data-quality techniques).

Additionally, there is the potential for business opportunities through the application of fees for data consumption (depending on queries, rates and so forth).

Toward a federation of data spaces

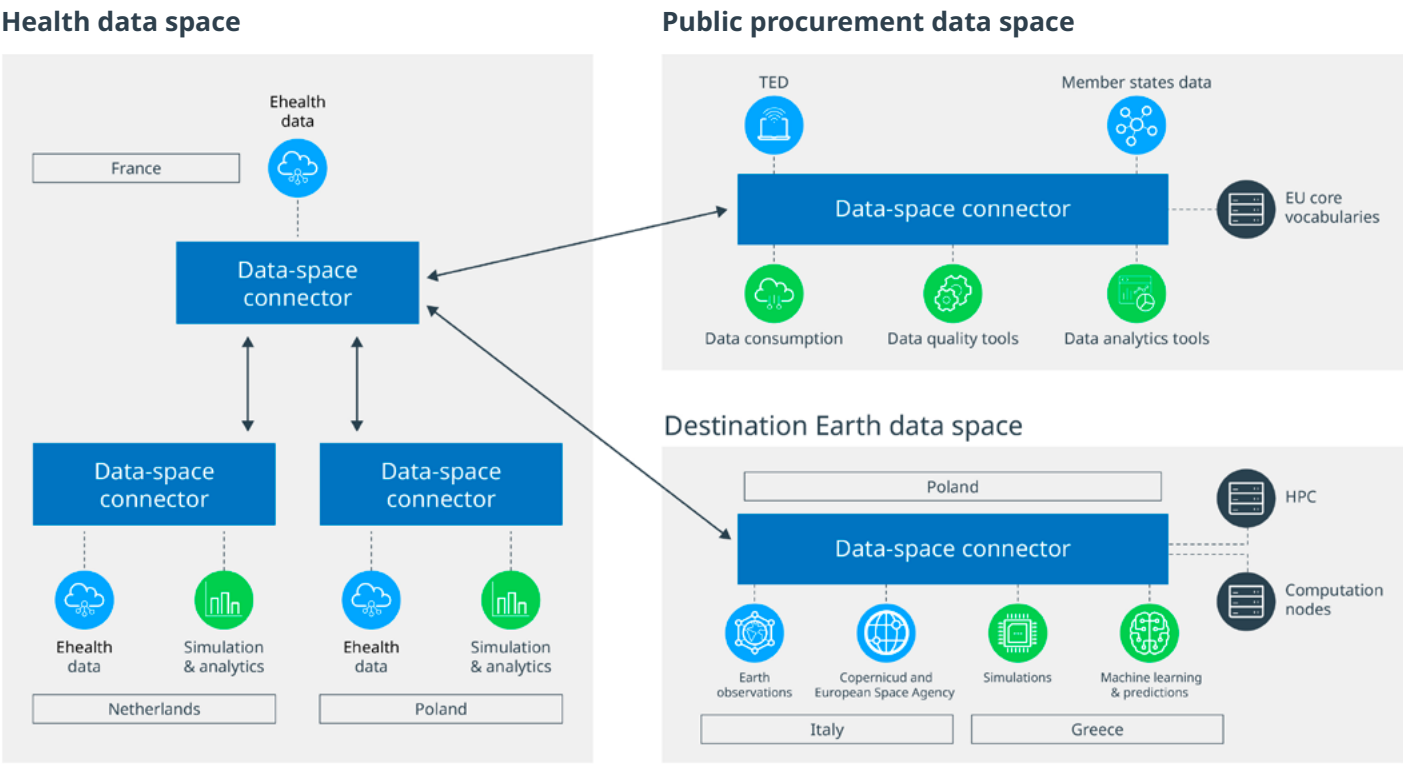


Figure 1: Example of a health data space

Of course, these entities will be required to seek authorization to access data sets, and adhere to governance rules.

As an example, under proper governance and consent, data could be shared between an electronic health data space and Destination Earth data space. These combined data sets would help determine the effect of the climate crisis.

Examples of data spaces in the public sector

By examining specific use cases in domains such as public procurement, mobility, health and skills, we can better understand how data spaces are being deployed to tackle persistent issues and drive innovation in the delivery of digital public services.

From consolidating and harmonizing public procurement data to enhancing mobility planning and management, and leveraging genomics in healthcare, the use cases presented in this section demonstrate the transformative potential of data spaces when the right foundations are in place.

The Public Procurement Data Space (PPDS)

Public procurement, vital to the EU's economic vitality, suffers from inconsistent data practices across administrations. This lack of harmonization, which results from disparate data sources, varied quality of data and the absence of common access standards, complicates evidence-based decision-making.

The Public Procurement Data Space (PPDS) aims to consolidate, harmonize and facilitate access to public-procurement data. It was developed in response to the need for a unified platform that provides easy access to comprehensive information on public-procurement markets under EU regulations.

By providing access to standardized and detailed data, the PPDS supports the analysis of information to improve decision-making, thus contributing to more informed and effective public policies. Moreover, it facilitates access to market information and procurement procedures conducted by other institutions, and it promotes quantitative and empirical research.

Future developments of the PPDS will involve connections to other data spaces that will support the generation of new use cases. For example, linking data with the European Data Space for Smart Communities will improve the operational management of the procurement of works, goods and services for cities, based on actual city needs, determined by data collected from Internet of Things (IoT) devices.

As the PPDS evolves alongside other data spaces it will not only transform the landscape of public procurement data across Europe but also serve as a cornerstone for more transparent, efficient and responsive public governance.

The European Mobility Data Space (EMDS)

The European Mobility Data Space (EMDS) is a fundamental part of the EU's Data Sharing and Interoperability Strategy. Nowadays, a lack of data homogeneity, disparity in semantics and divergences in data quality make it difficult to officially confirm and take advantage of data from different actors in the mobility sector.

The EMDS, together with the rest of the initiatives based on the European Data Strategy, should allow cities, regions and different types of transportation authorities to better plan their multimodal systems, routes, frequencies, quality of services and, of course, improve operational efficiency. To this end, the data used should be harmonized so it can be shared. Several rules of use should be established to ensure privacy while also facilitating the provision and consumption of data.

The EMDS model must facilitate the decentralized exchange of data among participants without duplicating information and still allow users to continue to be owners and managers of their data as well as consumers of third-party information.

The DeployEMDS project

NTT DATA participates in the DeployEMDS project, which is cofunded under the EU's Digital Europe Programme with the aim of making a common European mobility data space a reality.

The initiative will cultivate a broad European ecosystem of data providers and users, facilitating the adoption of common building blocks and validating them in 16 use cases from 9 EU countries.

The EMDS will offer a framework for interlinking and federating ecosystems. DeployEMDS supports the EMDS initiative through real-life implementation projects, a common technical infrastructure and common governance mechanisms.

This project and other similar initiatives are endorsing mobility management in support of a data-driven decision-making culture where decentralization must facilitate sharing.

Genomics in the European Health Data Space

This use case presents the combination of an existing data space with a powerful analytical platform and techniques.

High-performance genomics technologies (such as genome sequencing) integrated with patient data facilitate advances in disease diagnosis, management and prognosis. Accessible genomic and clinical data is essential for biomedical research and personalized medicine.

In this perspective, we outline the organization of genomic study platforms along three axes, aimed at providing analytical and data-governance capabilities linked to the European Health Data Space (EHDS) and national and regional data spaces:

1. **Regional genomic platform**
   The first axis involves a regional genomic platform supporting data storage, analysis and use across healthcare regions. This centralized system optimizes resources and enhances clinical practice and research management. It fosters unified, accessible databases and facilitates informed, preventive clinical decision-making, while supporting secondary uses such as research and health-policy development.
2. **National genomics and health data platform**
   The second axis comprises a national genomics and health data platform, managed by each European country's governmental health authority. The platform facilitates both the primary and secondary use of healthcare data and ensures standardized collection and organization nationwide. This unified database promotes cross-regional collaboration, mitigates interoperability issues and offers accessibility to regions without their own genomic platforms.
3. **Integration of national genomics platforms**
   The third axis acts as a connector within the EHDS ecosystem, linking each country's national genomics platform. This integration is crucial for improving healthcare delivery and advancing research across Europe. Overcoming challenges such as interoperability and quality assurance requires centralizing, standardizing and ensuring access to genomic information across EU member states. This connection supports collaboration and data exchange within the EHDS framework while promoting cross-border projects among member states.

With the EHDS, new services can support researchers, policymakers and health professionals in their activities while ensuring that high-quality data can be exchanged with appropriate governance in place.

The European Data Space for skills

Skills are both highly personal for citizens and omnipresent in myriad policies and services related to work, education, migration, entrepreneurship and more. In other words, skills are an essential part of the European economy.

As the aim of common European data spaces is to make more data available for the benefit of European organizations and citizens, the value and potential of skills in the context of these data spaces cannot be overstated.

Practical use cases for a skills data space are easy to find. Sharing data on skills demand (such as job vacancies) allows citizens to access a wider range of opportunities, while data on skills supply (such as CVs) gives employers a larger pool of potential candidates. In the context of skills shortages in different sectors, such an exchange offers tremendous economic value.

In addition, making skills data widely available allows for data analysis on a wider, more complete dataset. This, in turn, results in more accurate forecasting to support policy- and decision-making.

The European Data Space for Skills (DS4Skills) project aims to achieve precisely that. Central to such a data space is interoperability to facilitate data exchange, illustrated by the use cases described above.

Considerations regarding data spaces in the public sector

The successful implementation of data spaces requires tackling a range of complex issues: data, governance, digital services and sovereignty/trust. By exploring these considerations in depth, public-sector leaders can develop a more comprehensive understanding of the complexities involved and identify strategies to overcome them.

Data

Data is the central asset of a data space, which focuses on the value of the data and the potential economy around the data that is shared, exchanged and manipulated in the data-space ecosystem.

Several critical challenges appear during the implementation and operationalization of the data space.

Data quality

The main challenge lies in effectively exposing and sharing reliable data that permits the identification of registries or facilitates the sharing of registries with sufficient information and corresponding metadata that clearly describes the data.

Inconsistencies between information and metadata in datasets may cause issues when data from different sources is integrated.

Let's suppose that we have an employee dataset containing several atomic data elements. We may need to integrate different data sources that have inconsistent datasets and inconsistent data labels (naming the variables in the dataset differently). Integrating datasets and extracting knowledge is a tedious exercise that requires thorough data cleansing and enrichment.

Data spaces should therefore implement mechanisms to deal with such inconsistencies. To achieve this, a finer-grained approach is necessary, involving atomic data (such as "name" in our employee dataset).

Data security and privacy

Another key challenge is protecting sensitive information from unauthorized access and ensuring compliance with data-protection regulations.

Raw or original data must be managed and maintained at the source (origin) to ensure correct management by the data holder, especially when the data is personal.

Only metadata or processed data from the original (for example, the result of aggregations or trained algorithms) can be transferred to external entities. This creates privacy and security concerns relating to the data spaces themselves and the corresponding data management.

Data interoperability and integration

The main challenge in this regard is the integration of heterogeneous and diverse information from different use cases and domains. The data ecosystem inside a data space complies with different formats, structures, standards and legal requirements (and definitions) that should be harmonized and made interoperable to support data-driven decision-making and policymaking.

Modernizing legacy systems toward application and infrastructure interoperability aligns with the interests of the EU by empowering data computation and exploitation inside the EC boundaries as a base for a future economy based on data.

Governance

Data-governance frameworks

One of the main challenges in data-space adoption is the design, development and maintenance of a data-governance framework inside a public administration as a basis for the long-term sustainability of data spaces.

Such a framework should be clear about the processes and responsibilities concerning data-space management, from the management of data to applications.

Thus, there is a need for multiple (digital) public services to collaborate under data spaces, interrelating different administration bodies.

Identification of roles inside the data space

Multiple roles with different responsibilities are needed to operationalize, maintain and scale up data spaces. The challenge here lies in creating a community that self-governs data spaces as a team composed of different companies with different business objectives.

Capacity building

A third challenge is the generation of skills inside and outside public administrations to manage and govern data spaces effectively.

Digital services

Digital services that will act as either consumers or providers of the data space are the most relevant actors facilitating the adoption of data spaces and extracting value from the data.

Create a data culture and maximize the adoption of applications

Empower both consumers and providers to use the data space. Efforts should focus on cultivating an understanding of data-space roles and emphasizing the importance of data sharing and the long-term benefits of this new paradigm.

Maximize collaboration between organizations and stakeholders

This will enrich the data, its context and even the generation of applications under a common framework. Data spaces should facilitate the establishment of business and economic structures around the data lifecycle through data-centricity.

Sovereignty and trust

Data sovereignty is of central importance to the success of data spaces and the future of the data economy – and it poses specific challenges. Each data owner should have the full right of determination regarding the use of their data. This implies that owned data needs to be traced.

To control how data is used, one must first and foremost know about its use. Because data spaces are nested and connected, this can become complex. It requires specific technologies (such as data-space connectors, marketplaces, distributed architectures and edge computing) to maintain this control.

Control over data is also a technically challenging problem: for instance, how do we ensure that the use of data can be disallowed once it is used in a complex data ecosystem?

Some of these problems might be intractable, which means that we will have to find compromises when drawing up the smart agreements governing data spaces.

High-level data-space architecture

Once the core problem to be addressed and the specific use cases for data sharing are defined, organizations can then focus on understanding the diverse roles, governance frameworks and technical building blocks that come together to bring data spaces to life.

By gaining clarity on core concepts – key stakeholders and their responsibilities, the trust frameworks and protocols that govern data exchange, and the connector technologies that facilitate interoperability – public-sector leaders can develop a comprehensive understanding of the ecosystem required to sustain impactful data spaces.

Drawing on NTT DATA and AWS's extensive experience, this section provides practical guidance on hosting, securing and architecting data spaces on the AWS cloud.

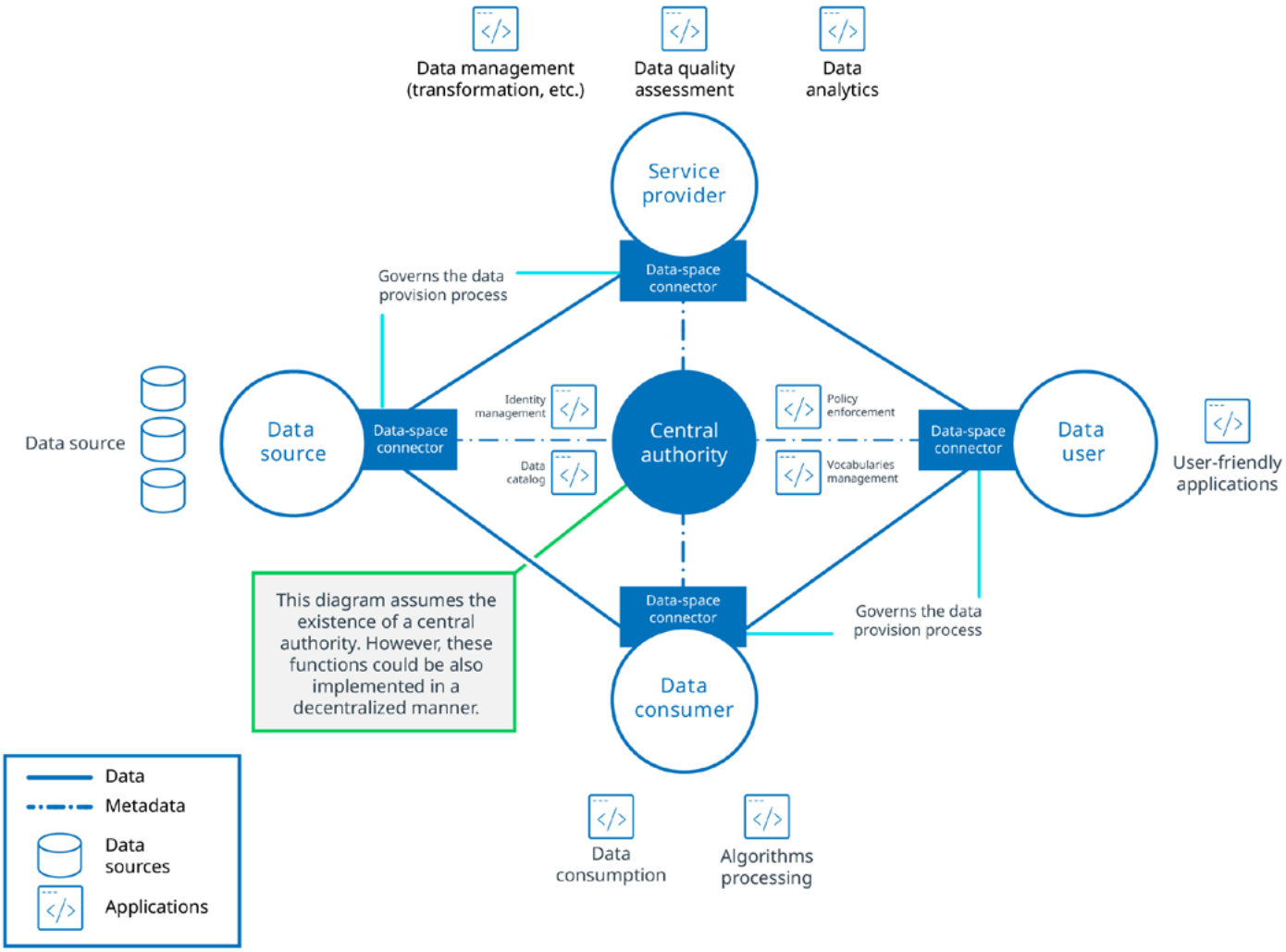


Figure 2: Roles in data spaces. Source: NTT DATA

Understanding roles in data spaces

Data producers

Data is created by data producers, which can be IoT devices, computer applications or people. They are not necessarily the data owners. Who owns the data and what data ownership encompasses are important considerations and subject to a set of policies.

Data providers

Data providers are organizations or individuals that have data which can be consumed by the integrating entities of the data space.

Data processors

Data processors are the parties or organizations entrusted by owners, producers and users to provide services and act on data. Data actions include storage, mergers, authentication checks and authorization. As the intermediary between the producers, owners and users, data processors allow or request access or actions.

Data-space governance bodies

It is assumed that each data space is managed by a body that establishes governance rules for it – both access rules (authentication and authorization) and interoperability-related rules (for example, on metadata).

Data processors and data-space governance bodies may be the same entity, but not necessarily. Data processors can also provide technologies, software and infrastructure for acting on the data on behalf of one or several data-space governance bodies. They are also expected to create a bridge between separated data spaces by mechanisms of interoperability, and to provide data traceability (data lineage) services and technologies.

Data infrastructure providers

These providers deliver the infrastructure (mainly cloud, edge and edge-to-cloud infrastructure and networks). They also host applications that perform data actions on behalf of data processors and may play the role of data processors (in full or in part).

Data consumers

Data consumers are organizations or users who use the data they receive from data providers or any services in the data space provided by other participants, respecting the usage policies that correspond to their profiles.

Data marketplace agents or services

This role can also belong to data processors.
Data marketplace agents or services establish and run mechanisms to provide economic benefits from the use of data to data users, data owners and other actors.

A general EU data-spaces governance body should exist to rule on the interactions between data spaces and, importantly, to define interoperability rules.

Common building blocks for a data space

App store

An app store securely distributes data apps that facilitate data-processing workflows. It allows searching for apps within the marketplace, and apps can be registered, published and maintained. The store provides the app metadata, and publishing, using and finding these apps is done through standardized processes as explained by the International Data Spaces (IDS) Association.

Metadata broker

A broker service provider manages information about the data sources available in the data space. It registers metadata about the data offers and conditions for using that data, serving as a catalogue.

Data sinks

Data sinks are applications that execute on the consumer side to process, transform or store data.

Connectors

Connectors are software components that enable data to be shared and integrated between sources.

Figure 3 illustrates how these services work together, as defined by the International Data Spaces (IDS) Association.

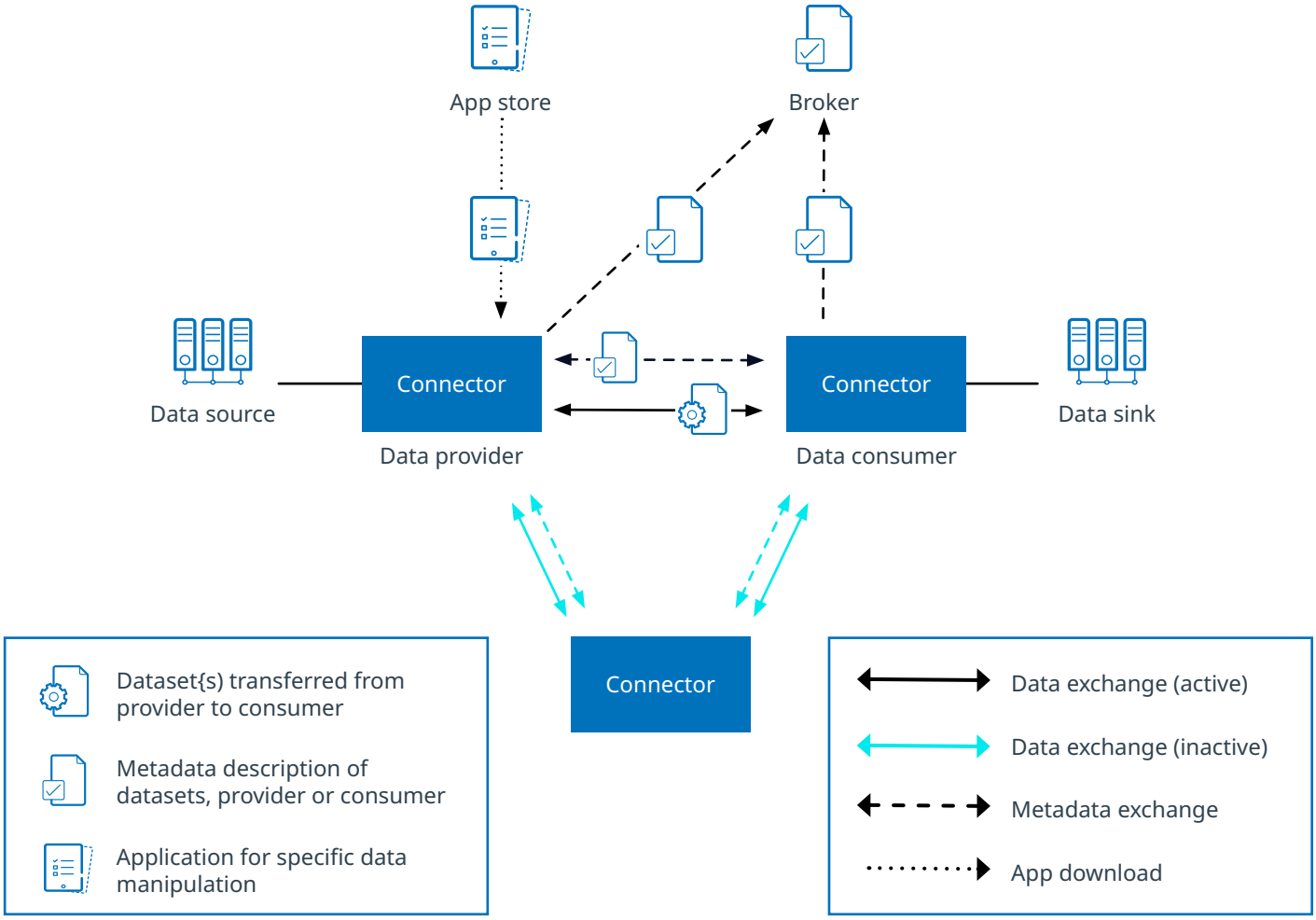


Figure 3: Interactions between technical components as depicted in the International Data Spaces (IDS) Reference Architecture Model

Trust frameworks

Trust frameworks define a set of guidelines, standards and principles for the trust and security measures in a data space. They outline the rules, policies and best practices for the secure exchange of data among participants.
The Gaia-X and IDS Trust frameworks are commonly used.

The **IDS Trust Framework** provides a solution for data exchange between organizations and individuals by enabling secure and efficient data sharing, processing and use. It describes the functional requirements (mandatory and optional) required to build trusted data spaces in centralized, federated or decentralized architectures.

The **Gaia-X Trust Framework** addresses data sovereignty and interoperability challenges by ensuring that organizations retain control over their data. It defines a set of rules and the minimum requirements to be part of the Gaia-X ecosystem (which provides common governance and a basic level of interoperability).

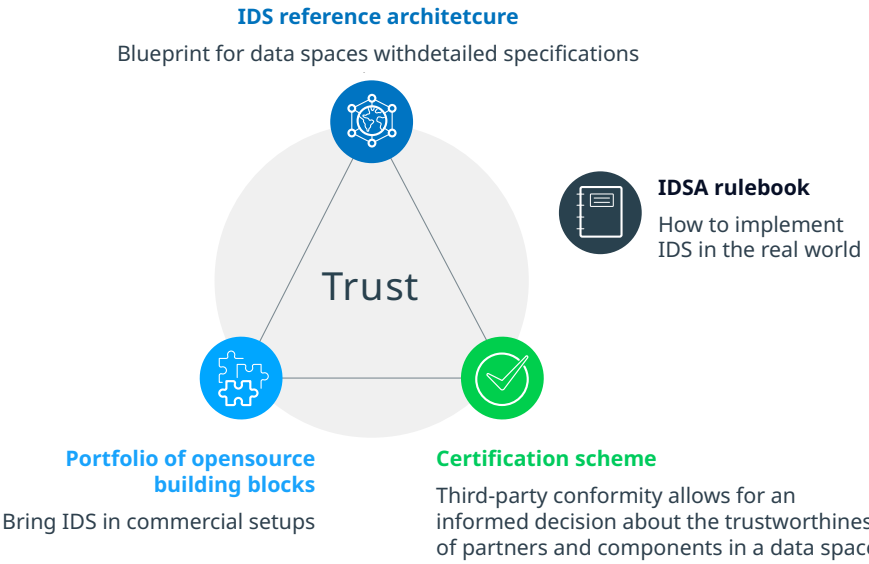


Figure 4: Interactions between technical components as depicted in the IDS Reference Architecture Model

The Dataspace Protocol

The Dataspace Protocol defines technical guidelines and rules governing the interchange, access and transmission of data in a data space – including data exchange, data-ownership preservation, interoperability and effective participant communication.

It defines key concepts and components that act as the basis for standardization and interoperability of data exchange:

- **Data representation and cataloging:** This defines the structure and format of the data being shared.
- **Data assets:** These constitute the pieces of data published to a data space. They can be versioned, and their metadata can include information such as timestamps, authors and descriptions.
- **Data services:** This functionality is provided by a data space to perform operations on assets, such as querying, filtering or transforming data. Services can be invoked using REST application programming interfaces or message queues.
- **Exchange policies:** These rules dictate how data can be accessed, modified or deleted. Data-usage and data-control policies can be defined at multiple levels, including at the organizational, dataset or asset level. The policies are attached to each asset through a connector. Policy violations can initiate alerts and actions to enforce data governance in the data space.

Connector technologies for data spaces

Connector technologies implement the connectors that provide and manage the data-space policies, also called the data-exchange rules.

The most relevant connector technologies are the Eclipse Dataspace Components framework, the FIWARE TRUE Connector and the Simple platform.

The **Eclipse Dataspace Components framework** has been developed by the Eclipse Foundation as free, open-source software. The goal is to create an efficient and functional data-transfer component that implements the protocols of the IDS standard and is compatible with the Gaia-X project. As a central component, the connector enables the exchange of data through defined data sovereignty contracts that are automatically negotiated to govern access to data assets.

With a focus on extensibility and adaptability, the architecture of the Eclipse Dataspace Components (EDC) was developed based on feedback from the IDS and Gaia-X initiatives. The EDC framework is designed and built on four pillars: identity, trust, sovereignty and interoperability.

The **FIWARE TRUE Connector** provides a specification that an organization can use to share data securely and efficiently in the IDS ecosystem. It provides a standardized way of exchanging data securely and in a traceable manner. The tool consists of three main components: the execution core container, the data application and the usage-control data application. These components work together to enable data exchange, communication with identity providers and enforcement of usage-control policies.

Using FIWARE TRUE Connector, an organization can benefit from secure, efficient and interoperable data sharing. The latest version of the connector is implemented using the technical recommendations provided in the Dataspace Business Alliance's Technical Convergence document.

Simpl is a smart middleware platform that represents a significant step toward common European data spaces. It addresses the challenge of resource sharing while preserving control and security, which fosters trust among stakeholders, making it a promising solution for public- and private-sector entities.

Collaboration is essential, and Simpl acts as a common "glue", ensuring interoperability across diverse capacities without costly interfaces.

Data-space connectors are an essential component of data spaces, and NTT DATA and AWS have experience both as contributors to their development and as users.

For an in-depth comparison between different connector technologies, refer to the IDSA's Data Connector Report.

Data spaces on AWS

AWS has played a pivotal role in shaping data spaces and collaborative ecosystems across industries. By providing a robust and scalable global infrastructure, and with over 200 managed services to choose from, AWS has empowered organizations to create and manage data spaces that facilitate data sharing, collaboration and innovation.

Hosting a data space on AWS

AWS offers different options for hosting your data space, including Regions, Availability Zones, AWS Local Zones and AWS Outposts. As the client, your first option is to select a Region of your choice based on your preferences as well as sovereignty constraints. AWS offers eight Regions across the EU, each containing three availability zones for increased availability and fault tolerance.

Once a Region is selected, you need to create a Virtual Private Cloud (VPC) to deploy the Data Space frameworks and services. A VPC lets you logically isolate your virtual network.

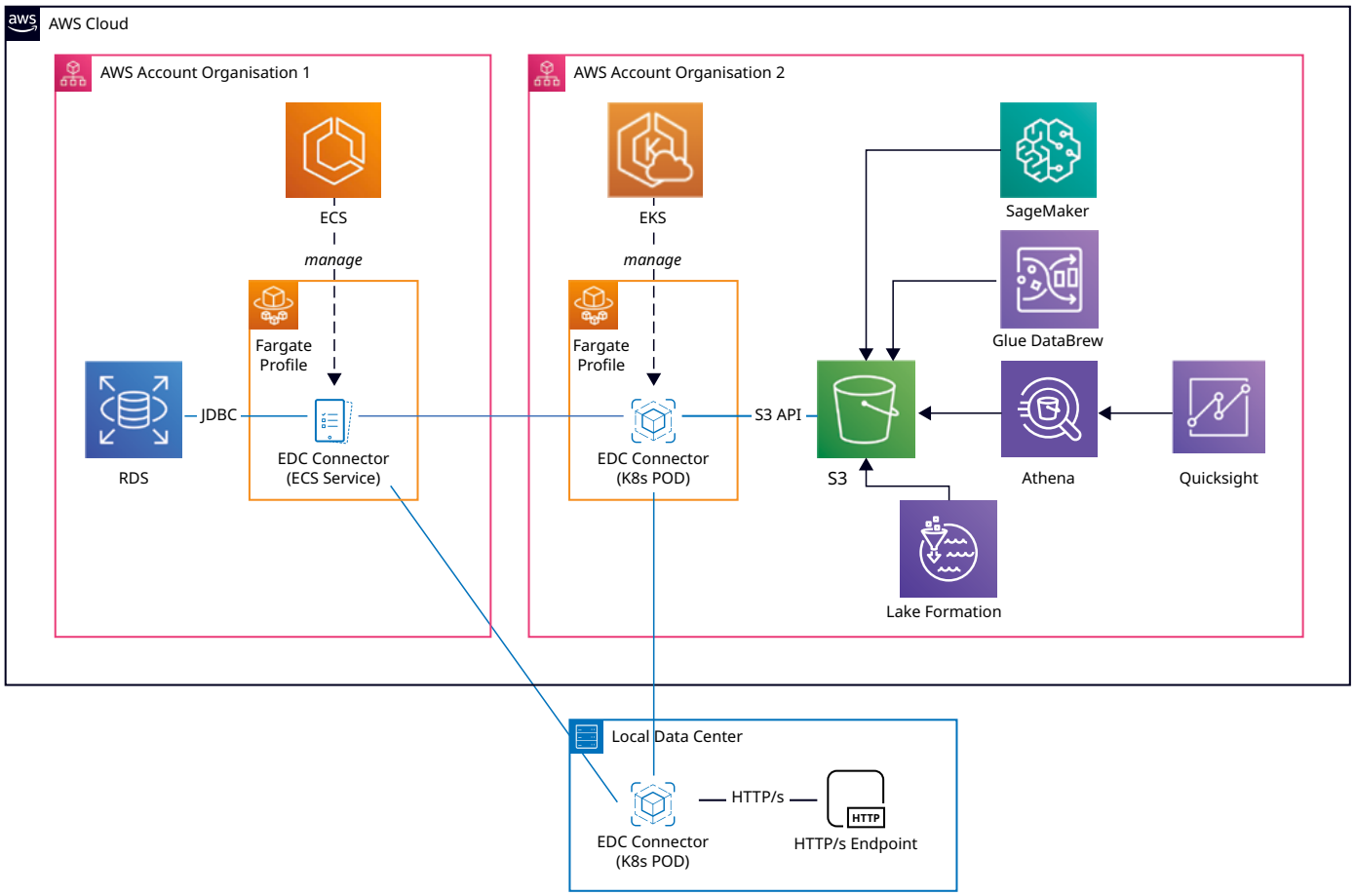


Figure 5: Sample data space architecture on AWS . Source: AWS

Data security in your virtual private cloud

You have complete control over your virtual networking environment in your VPC, including the selection of your Internet Protocol address ranges, the creation of subnets, and the configuration of route tables and network gateways.

Security is a shared responsibility between the client and AWS. The client is responsible for maintaining control over the content hosted on AWS infrastructure. There are different components for securing your data in a VPC. Some recommendations to secure your data include using multifactor authentication, using SSL/TLS, logging user activity (with AWS CloudTrail) and using AWS security controls.

Architecture of a data space on AWS

After choosing the Region and creating the VPC, we are now ready to deploy the different services. In this architecture, the data-space connector is deployed on docker containers by using Amazon Elastic Container Service (ECS) with AWS Fargate or Amazon Elastic Kubernetes Service (to run Kubernetes on AWS). To debug and monitor distributed applications, use AWS X-Ray and AWS App Mesh.

The data from the data space will reside in Amazon Simple Storage Service (S3) and Amazon Relational Database Service (RDS), which are used as data sources/sinks. To extract meaning from data by processing it, use AWS Glue to build extract, transform and load pipelines.

Once you have deployed your solution, you need to monitor it (using Amazon CloudWatch or AWS CloudTrail).

By using these managed and serverless technologies, you can build data spaces that are scalable, reliable and cost-effective. You can also focus on building applications and services that can take advantage of the power of data, without having to worry about the underlying infrastructure.

Extending a data space on AWS

In addition to the foundational building blocks for the Data space, AWS provides services that augment a data space's functionality, such as Artificial Intelligence and Machine Learning (Amazon SageMaker), Real-time Analytics (Amazon Redshift, Amazon Athena), Data Visualization (Amazon QuickSight), and much more.

Organizations can leverage these services as building blocks and provide further benefits to end customers.

AWS and NTT DATA: value proposition and lessons learned

AWS and NTT DATA leverage their respective strengths to support organizations navigating the complexities of data spaces, where seamless integration, scalability and advanced analytics are paramount. They enable public- and private-sector organizations to create and manage impactful data spaces.

NTT DATA

NTT DATA plays a pivotal role in shaping strategic models for data-oriented public services. Our commitment to enhancing interoperability and promoting the use of shared data is reflected in our contributions to initiatives such as the PPDS and the EMDS.

NTT DATA assists public administrations in digital transformation through analytics and data-mining projects, defining strategic models tailored to the ecosystem's reality. Notable implementations of data spaces include the Once Only Technical System (OOTS) and interconnections within the CyclOps European project. NTT DATA provides advisory services to European initiatives like Gaia-X, IDSA and BDVA, fostering interoperability across digital public services with frameworks like the European Interoperability Reference Architecture (EIRA) and Common Assessment Method for Standards and Specifications (CAMSS).

NTT DATA's advisory role extends to projects for entities like the European Union Agency for Railways and the Spanish Ministry of Justice, facilitating the evaluation of public policies in future justice data spaces.

Amazon Web Services

In tandem, AWS plays a pivotal role in supporting data spaces and collaboration across industries, in line with the company's commitment to support organizations in using data for innovation and addressing global challenges. AWS's role as a trusted infrastructure provider ensures that clients have access to a reliable and secure environment for managing their data spaces.

AWS also offers prescriptive guidance and reference architectures to streamline deployment processes, and its extensive portfolio of advanced analytics and AI and machine learning services empowers organizations to derive deeper insights and unlock opportunities within their data spaces. Notable projects, such as the Smart Freight Centre's SFC Exchange Network and Bosch's COVANTO platform, as well the company's involvement in initiatives like Gaia-X and Catena-X, underscore AWS's dedication to supporting collaborative data initiatives and driving innovation across industries.

Together, NTT DATA and AWS offer a holistic approach to data-space creation and management, enabling organizations to seamlessly ingest, store, analyze, and derive actionable insights from their data with unprecedented efficiency and agility. Their partnership facilitates secure, efficient and scalable data solutions that empower both public- and private-sector organizations.

By accelerating the adoption of digital services, this collaborative partnership drives innovation, accelerates time to value and helps organizations succeed in the rapidly evolving data landscape.

10 key lessons

Drawing on their extensive experience in supporting government data initiatives, NTT DATA and AWS identified the following 10 lessons concerning the development of public data spaces:

1. Creating value through data sharing is central

The public data space needs to fulfil a purpose and bring value to participating government agencies and citizens. Value creation must be at the heart of the data-space design. Developing a compelling business case, supported by practical examples or case studies, enhances clarity and reinforces the value proposition to stakeholders. Conducting cost-benefit analyses, measuring key performance indicators and providing practical implementation strategies empower government organizations to prioritize resource allocation and align data-space development with their policy objectives.

2. Integrate data with governance

Integrating data across different government agencies is an initial and crucial step toward harmonizing and scaling up the integration process. This integration requires coordination among various government stakeholders and departments. Establishing clear and robust data governance mechanisms in the public data space facilitates the management of data, catalogs and metadata.

It's also essential to develop a robust framework that supports the coexistence of diverse data models and query languages, ensuring seamless convergence between different government data sources.

3. Ensure data integrity at the source

Maintaining data integrity and compliance with regulations is critical for building trust within the public data space. Emphasizing the importance of ethical data-handling practices and privacy-protection strategies reinforces government organizations' commitment to safeguarding sensitive citizen information. By safeguarding the integrity of data from its origin, government agencies uphold regulatory standards and protect sensitive information, fostering trust and accountability within the public data space.

4. Interoperability is critical

Interoperability extends beyond technical compatibility to include governance, legal and semantic aspects. Emphasizing stakeholder engagement and collaboration throughout the public data-space lifecycle facilitates the alignment of diverse government interests and priorities. Practical implementation strategies for addressing interoperability challenges, such as semantic mapping and standardization efforts, enable seamless data exchange and collaboration across heterogeneous government environments.

5. Go from micro data spaces to national data spaces

Establishing and managing micro data spaces in government agencies is a foundation for scaling up the technology, data-sharing strategies and more. A natural scale-up will facilitate more data sharing, trust and privacy mechanisms, or even mechanisms to facilitate the discovery and exploration of government information.

6. Scalability in data-space creation involves a multifaceted approach

Architecture design with dynamic scalability and flexibility is the cornerstone of overcoming limitations. This entails the adoption of scalable storage solutions, distributed computing frameworks and elastic infrastructure provisioning to allocate resources and adapt to evolving government requirements and workload demands. Using cloud computing, containerization and orchestration technologies further enhances scalability and agility, facilitating seamless scaling as demand increases.

Moreover, aligning the storage and representation of government information with policy needs is imperative. Organizing information appropriately enables the creation of efficient queries and facilitates information retrieval, which is particularly critical in the presence of large government datasets. Balancing the scale-up of information with optimal performance keeps the public data space robust and responsive to the government's evolving needs.

7. Use reference architectures so you don't have to reinvent the wheel

Leveraging reference architectures such as EIRA and eGovERA to model and design data architectures for digital public services streamlines implementation and publication processes. These frameworks provide traceability from government requirements to solution selection and adoption, promoting efficiency and consistency.

It's crucial for government organizations to reuse as many modules as possible to optimize resource use. This approach also ensures compatibility with standards and accelerates time to market for data-driven government initiatives.

8. Invest in technical expertise and capacity building

Recognizing the technical complexity of public data spaces highlights the need for government organizations to invest in professionals who are adept in data integration, governance and cybersecurity.

Providing training and development opportunities bridges skill gaps and ensures the effective management of public data-space infrastructure. Moreover, through training, government organizations empower their workforce to use data effectively, foster a data-oriented culture and contribute to the success of the public data-space initiative.

9. Implement monitoring and evaluation for optimization

Implementing robust monitoring and evaluation mechanisms is vital for assessing public data-space performance and effectiveness.

By monitoring data use, access patterns and compliance metrics, government organizations can identify areas for improvement and inform decision-making processes. Integrating feedback loops fosters a culture of continuous learning and innovation, allowing for the iterative refinement of data governance policies and technology infrastructure.

This approach empowers government organizations to adapt and evolve, maximizing the value derived from public data-space investments and supporting sustainable growth and innovation.

10. Focus on community engagement and collaboration

Engaging with the broader community of stakeholders, including other government agencies, academic institutions and civil society organizations, fosters collaboration and the cocreation of value in public data spaces.

In the case of government data spaces, a vibrant community of policymakers, academics and other stakeholders should be set up to codesign the data-space approach and use cases. Special attention should be paid to methods to ensure data quality – a major challenge in the public sector.

The vision for data spaces in public administration

Data spaces are emerging as a powerful tool for data sharing, integration and collaboration across the public sector. By harnessing their potential, public-sector entities can improve the quality and efficiency of service delivery, enable data-driven policymaking and foster new value-added services.

Looking ahead, the vision for data spaces in public administration involves the transition from micro-level initiatives to pan-European, interconnected platforms. This scale-up will be crucial for realizing the full potential of data spaces, empowering public sectors to be more transparent, make better-informed decisions and dynamically adapt to evolving challenges.

Ultimately, data spaces hold the key to a digitally transformed future where public administrations can more effectively serve their citizens and address complex global needs.

By embracing this new paradigm and leveraging the collaborative strengths of partnerships like that between NTT DATA and AWS, the public sector can unlock unprecedented value from its data assets and drive meaningful progress for society.

Take the next step

Are you ready to unlock the power of data spaces in your public-sector organization? NTT DATA and AWS stand ready to support you on this journey. By combining our deep expertise in data-driven transformation, cloud infrastructure and data governance, we can help you:

- Assess the readiness and potential impact of data spaces within your organization.
- Develop a comprehensive strategy and roadmap for data-space implementation.
- Design and deploy scalable, secure and interoperable data-space architectures.
- Foster the necessary organizational capabilities and data-driven culture.
- Continuously optimize and evolve your data-space initiatives for maximum value.

Take the first step today by connecting with NTT DATA and AWS experts to explore how data spaces can drive your digital government initiatives forward.

Together, we can navigate the complexities and harness the full transformative power of data to better serve citizens and unlock new frontiers of innovation.

Contact us

References

- Amazon Web Services. (2019). Integrating AWS X-Ray with AWS App Mesh. AWS Compute Blog. Retrieved from https://aws.amazon.com/es/blogs/compute/integrating-aws-x-ray-with-aws-app-mesh/
- Amazon Web Services. (2023). AWS for data spaces. AWS. Retrieved from https://aws.amazon.com/es/government-education/aws-for-data-spaces/
- Amazon Web Services. (2023). Enabling data sharing through data spaces on AWS. AWS Public Sector Blog. Retrieved from https://aws.amazon.com/es/blogs/publicsector/enabling-data-sharing-through-data-spaces-aws/
- Amazon Web Services. (2023). Introduction to data spaces. AWS Prescriptive Guidance. Retrieved from https://docs.aws.amazon.com/prescriptive-guidance/latest/strategy-building-data-spaces/building-data-spaces.html
- Big Data Value Association. (2023). DEP data spaces. BDVA. Retrieved from https://bdva.eu/dep-data-spaces/
- Eclipse Foundation. (2023). Eclipse Data Space Connector. GitHub. Retrieved from https://github.com/eclipse-edc
- European Commission. (2022). Preparatory work in view of the procurement of an open source cloud-to-edge middleware platform - Architecture Vision Document. Retrieved from https://ec.europa.eu/newsroom/dae/redirection/document/86241
- European Commission. (2023). Simpl: Secure interoperability for linked data spaces. Digital Strategy. Retrieved from https://digital-strategy.ec.europa.eu/en/policies/simpl
- FIWARE Foundation. (2021). FIWARE for data spaces. FIWARE Position Paper. Retrieved from https://www.fiware.org/wp-content/uploads/FF_PositionPaper_FIWARE4DataSpaces.pdf
- Gaia-X. (2022). Policy Rules Document v22.04. Gaia-X. Retrieved from https://gaia-x.eu/wp-content/uploads/2022/05/Gaia-X_Policy-Rules_Document_v22.04_Final.pdf
- Gaia-X. (2023). Trust framework. Gaia-X Policy Rules Committee. Retrieved from https://gaia-x.gitlab.io/policy-rules-committee/trust-framework/trust_anchors/
- International Data Spaces Association. (2024). Data space protocol overview. IDS Knowledge Base. Retrieved from https://docs.internationaldataspaces.org/ids-knowledgebase/v/dataspace-protocol/overview/readme
- International Data Spaces Association. (2024). IDSA rulebook. IDS Knowledge Base. Retrieved from https://docs.internationaldataspaces.org/ids-knowledgebase/v/idsa-rulebook/idsa-rulebook/3_functional_requirements#technicalcomponents-of-a-dataspace
- Nordic Institute for Interoperability Solutions. (2023). Is X-Road a data space technology? NIIS Blog. Retrieved from https://www.niis.org/blog/2023/6/21/is-x-road-a-data-space-technology
- NTT DATA. (2022). NTT DATA publishes white paper on the architectural concept for its global trusted data spaces. NTT Data News. Retrieved from https://www.nttdata.com/global/en/news/press-release/2022/june/ntt-data-publishes-white-paper-on-the-architectural- concept-for-its-global-trusted-dataspaces
- NTT DATA. (2023). El impacto de los espacio de datos en la creación de servicios digitales para el sector público. NTT DATA España Blog. Retrieved from https://es.nttdata.com/insights/blog/el-impacto-de-los-espacio-de-datos-en-la-creacion-de-servicios-digitales-para-el- sector-publico
- NTT DATA. (2023). Trends in data space technology. NTT DATA. Retrieved from https://www.nttdata.com/global/en/insights/focus/trends-in-data-space-technology
- NTT DATA. (2024). The European Health Data Center. NTT DATA UK. Retrieved from https://uk.nttdata.com/insights/whitepapers/the-european-health-data-center
- Think IT. (2023). AWS pattern for multi-tenant data spaces. Think IT Lab. Retrieved from https://think-it.io/lab/aws-pattern-mvds


NTT DATA