COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI

MASTER UNIVERSITARIO EN INGENIERÍA DE TELECOMUNICACIONES

TRABAJO FIN DE MASTER

Feasibility study on the interconnection of data spaces across industrial sectors

Autor: Ignacio Núñez Gómez

Director: Atilano Ramiro Fernández-Pacheco Sánchez Migallón

Co-Director: Diego Mallada Conte

Madrid

2025

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título

Feasibility study on the interconnection of data spaces across industrial sectors

en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el

curso académico 2024/25 es de mi autoría, original e inédito y

no ha sido presentado con anterioridad a otros efectos.

El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido

tomada de otros documentos está debidamente referenciada.

Fdo.: Ignacio Núñez Gómez Fecha: 21/05/2025

Autorizada la entrega del proyecto

LOS DIRECTORES DEL PROYECTO

71216314B ATILANO RAMIRO FERNÁNDEZ-PACHECO

Digitally signed by 71216314B ATILANO RAMIRO FERNÁNDEZ-PACHECO
Date: 2025.05.21 15:48:07 +02'00'

Fdo.: Atilano Ramiro Fernández-Pacheco Sánchez-Migallón Fecha: ../../..

Fdo.: Diego Mallada Conte Fecha: ../../..

COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI

MASTER UNIVERSITARIO EN INGENIERÍA DE TELECOMUNICACIONES

TRABAJO FIN DE MASTER

Feasibility study on the interconnection of data spaces across industrial sectors

Autor: Ignacio Núñez Gómez

Director: Atilano Ramiro Fernández-Pacheco Sánchez Migallón

Co-Director: Diego Mallada Conte

Madrid

2025

Agradecimientos

A mi familia, por su apoyo incondicional a lo largo de estos 6 años de carrera.

A Diego, por haber contado conmigo como un miembro más de Gestamp desde el primer día.

A Atilano, por la ayuda y la confianza a lo largo de todo este curso.

FEASIBILITY STUDY ON THE INTERCONNECTION OF DATA SPACES ACROSS INDUSTRIAL SECTORS

Autor: Núñez Gómez, Ignacio
Director: Fernández-Pacheco Sánchez-Migallón, Atilano Ramiro
Co-Director: Mallada Conte, Diego
Entidad Colaboradora: Gestamp Servicios S.A.

RESUMEN DEL PROYECTO

En este proyecto se hace un estudio del estado actual de la tecnología de los espacios de datos y de su arquitectura. Se propone una red basada en las telecomunicaciones para habilitar una comunicación entre diferentes espacios de datos, y se trata de desplegar un conector para la integración de la entidad colaboradora, Gestamp, en Catena-X.

Palabras clave: Espacios de datos, Catena-X, Redes de telecomunicación, Conector

1. Introducción

El objetivo de este estudio es analizar la viabilidad de realizar interconexiones entre espacios de datos de distintos sectores industriales, utilizando a Catena-X como referencia, el principal espacio de datos del sector automotriz.

Este proyecto aborda diferentes etapas, que incluyen la investigación sobre la infraestructura de espacios de datos productivos, la propuesta de desarrollo de una red de espacios de datos similar al funcionamiento de Internet, y finalmente, el intento de integrar, a través de un EDC, a Gestamp en Catena-X.

2. Definición del Proyecto

Este proyecto se desarrolla en el ámbito de los espacios de datos, ecosistemas descentralizados cuyo objetivo es establecer marcos colaborativos entre sus miembros. Estos marcos suelen estar orientados a la transferencia de datos de forma estandarizada e interoperable, garantizando tanto la seguridad como la soberanía de la información.

Ante la creciente adopción de los espacios de datos en distintas industrias, surge la necesidad de crear una red interconectada de estos ecosistemas, que permita a las entidades intercambiar información no solo con empresas dentro de su propio espacio de datos, sino también con cualquier otra integrada en un espacio diferente. Esto resulta especialmente útil, ya que actualmente, si el departamento de ESG de una empresa desea transferir datos sobre emisiones de alcance 3 a dos compañías de espacios distintos, debe conectarse individualmente a los dos espacios de datos a los que pertenecen cada una de dichas compañías [1]. No existe aún un mecanismo que permita a una empresa permanecer conectada únicamente a su propio espacio de datos y, a través de un sistema de interconexión, comunicarse con otros espacios de manera directa.

Precisamente, este proyecto busca desarrollar ese mecanismo que actúe como puente entre distintos espacios de datos.

Sin embargo, conforme el proyecto ha avanzado, también lo ha hecho su enfoque. Inicialmente, la prioridad era diseñar un modelo para el intercambio de datos entre espacios. No obstante, el foco ha pasado a centrarse en el desarrollo de un conector que permita la integración de Gestamp en el ecosistema Catena-X.

3. Descripción del sistema planteado

Frente a la situación descrita, se propone un sistema de comunicación entre espacios de datos basado en infraestructuras de telecomunicación. Este sistema contempla la creación de un organismo dentro de Gaia-X que actúe de forma similar a los proveedores de servicios de Internet (ISP), redirigiendo el tráfico entre espacios de datos. Para ello, se asignarían BPNs públicas y privadas a cada entidad participante, permitiendo una comunicación segura dentro del ecosistema.

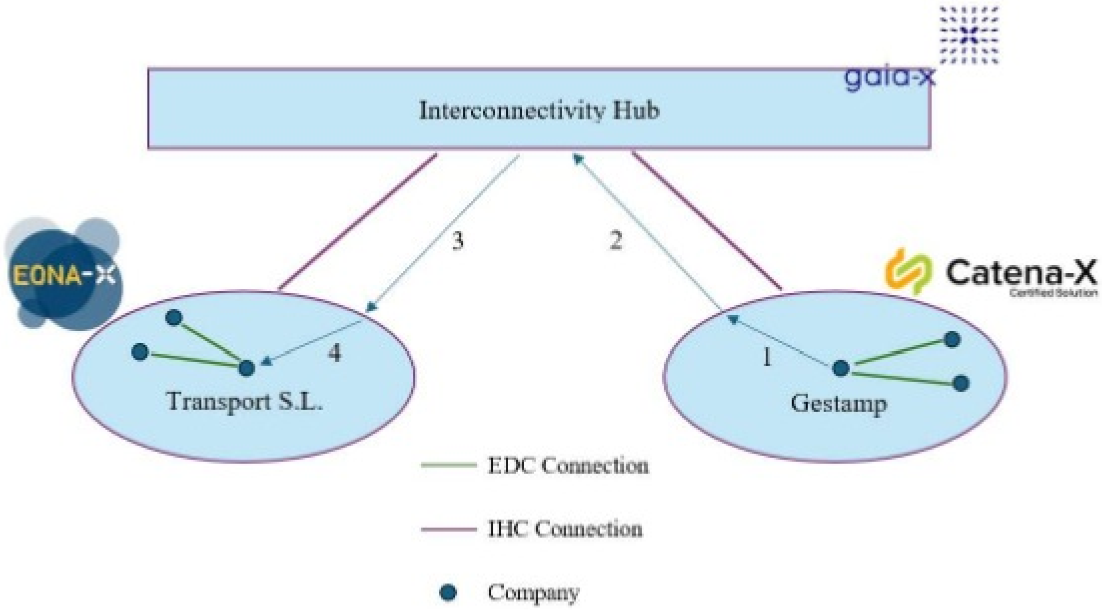


Ilustración 1. Propuesta de interconexión de espacios de datos a gran escala

4. Resultados

Dado que no se llegó a realizar una transferencia de datos entre espacios de datos, el resultado de esta fase se limita al análisis de los intentos de despliegue del EDC.

El despliegue del EDC se intentó mediante cinco métodos distintos. Los primeros cuatro enfoques buscaron desarrollar un EDC propio, pero no tuvieron éxito debido a que varios microservicios no se inicializaron correctamente, como se muestra en la Ilustración 2.

Ilustración 2. Estado de microservicios del despliegue del EDC propio

Sin embargo, el quinto método, que consiste en utilizar un EDC como servicio proporcionado por un proveedor externo, ha permitido que Gestamp se integre en Catena-X.

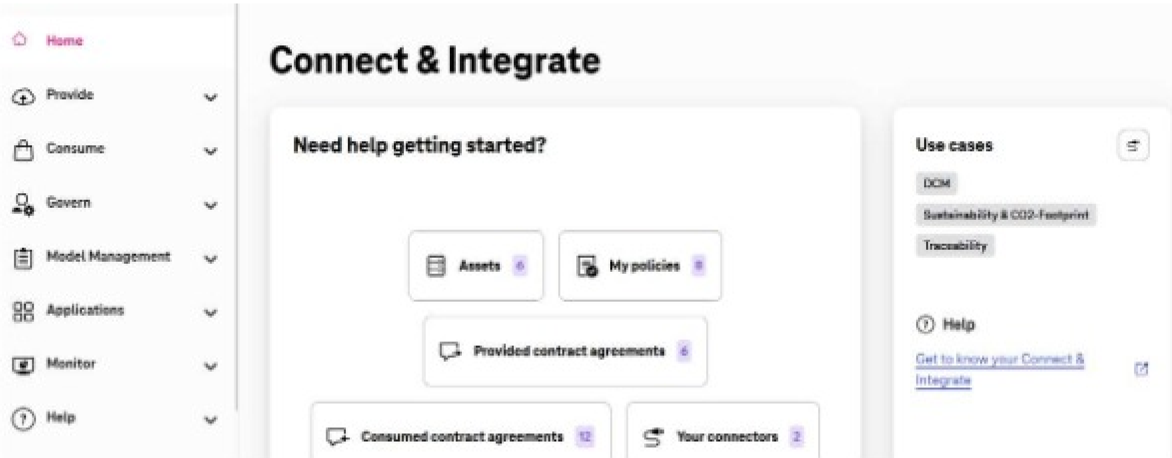


Ilustración 3. IU de la plataforma del EDC como servicio

5. Conclusiones

Aunque el enfoque inicial de este proyecto estaba dirigido a explorar la interconectividad entre espacios de datos, se dio un cambio de enfoque hacia un objetivo más fundamental: la integración y el despliegue del EDC en Cofinity-X.

Si bien es cierto que esto se ha logrado con éxito mediante el uso del conector como servicio, los siguientes pasos consisten en desarrollar un EDC local. Una vez se logre esto, se podrá proponer el desarrollo del IHC, el conector planteado para las comunicaciones entre espacios de datos.

6. Referencias

[1] EIT Urban Mobility. (2022). El potencial de los espacios de datos de movilidad (V1.0)

FEASIBILITY STUDY ON THE INTERCONNECTION OF DATA SPACES ACROSS INDUSTRIAL SECTORS

Author: Núñez Gómez, Ignacio
Supervisor: Fernández-Pacheco Sánchez Migallón, Atilano Ramiro
Co-supervisor: Mallada Conte, Diego
Collaborating Entity: Gestamp Servicios S.A.

ABSTRACT

This project studies the current state of data space technology and its architecture. A telecommunications-based network is proposed to enable communication between different data spaces, and efforts are made to deploy a connector for the integration of the collaborating entity, Gestamp, into Catena-X.

Keywords: Data spaces, Catena-X, Telecommunication Networks, Connector

1. Introduction

The goal of this study is to analyze the feasibility of making interconnections of data spaces across different industrial sectors, using Catena-X as a reference, the main data space of the automotive sector.

This project faces different stages, including research on productive data spaces' infrastructure, proposing the development of a network of data spaces similar to the way the internet works, and finally trying to integrate, via an EDC (Eclipse Dataspace Connector), Gestamp into Catena-X.

2. Project definition

This project is developed within the field of data spaces—decentralized ecosystems aimed at establishing collaborative frameworks among their members. These frameworks are generally focused on the standardized and interoperable exchange of data, ensuring both the security and sovereignty of information.

With the growing adoption of data spaces across various industries, there is a need to create an interconnected network of these ecosystems, enabling entities to exchange information not only with companies within their own data space but also with any other organization integrated into a different space. This is particularly useful because, at present, if a company's ESG department wishes to share Scope 3 emissions data with two companies from different data spaces, it must connect individually to each of those data spaces [1]. There is currently no mechanism that allows a company to remain connected solely to its own data space and, through an interconnection system, communicate directly with others.

This project aims to develop precisely that mechanism—a bridge between different data spaces.

However, as the project has progressed, so has its focus. Initially, the priority was to design a model for data exchange between spaces. However, the focus has shifted toward developing a connector that enables Gestamp's integration into the Catena-X ecosystem.

3. Proposed system description

In response to the described situation, a communication system between data spaces based on telecommunications infrastructures is proposed. This system involves the creation of an organization within Gaia-X that functions similarly to Internet Service Providers (ISPs), redirecting traffic between data spaces. To achieve this, public and private BPNs would be assigned to each participating entity, enabling secure communication within the ecosystem.

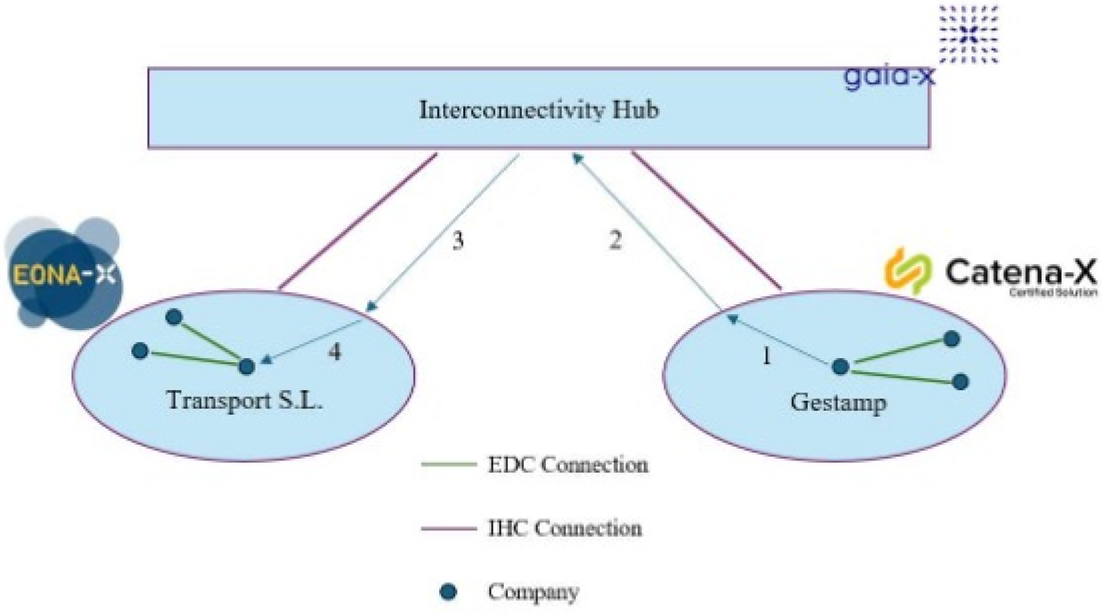


Ilustración 4. Data spaces interconnectivty proposal overview

4. Results

Since no data transfer between data spaces was ultimately carried out, the outcome of this phase is limited to the analysis of the EDC deployment attempts.

The deployment of the EDC was attempted through five different methods. The first four approaches aimed to develop an in-house EDC but were unsuccessful because several microservices were not initialized correctly, as shown in Ilustración 5.

Ilustración 5. Microservices status of the local EDC deployment

However, the fifth method, consisting in using an EDC as a service from an external provider, has enabled Gestamp to get integrated into Catena-X.

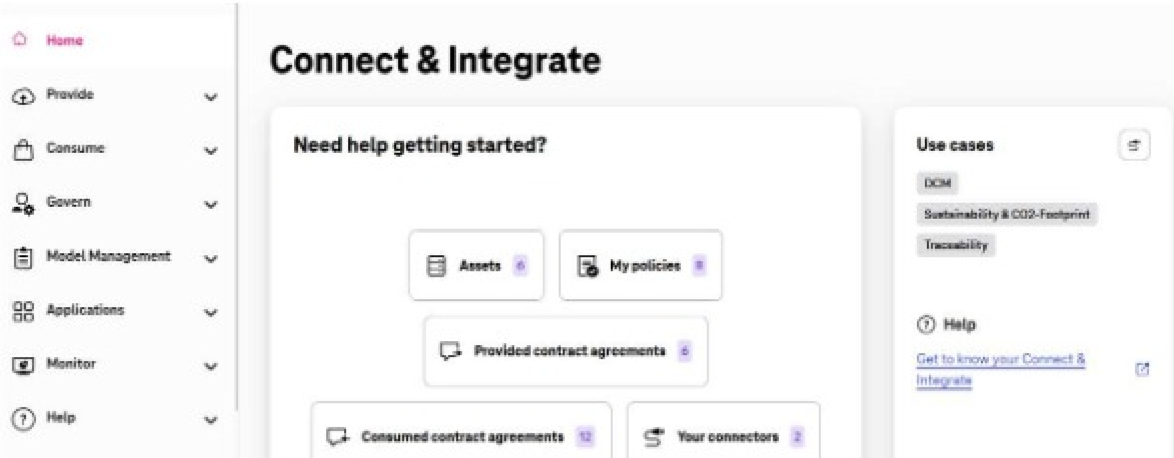


Ilustración 6. UI of the EDC as a service platform

5. Conclusions

Although the initial focus of this project was directed towards exploring interconnectivity between data spaces, there was a shift towards a more fundamental objective: the integration and deployment of the EDC in Cofinity-X.

While it is true that this has been successfully achieved using the connector as a service, the next steps involve developing a local EDC. Once this is accomplished, the development of the IHC, the proposed connector for communication between data spaces, can be proposed.

6. References

[1] EIT Urban Mobility. (2022). El potencial de los espacios de datos de movilidad (V1.0)

Index

Figure Index

Table Index

Chapter 1. INTRODUCTION

The goal of this study is to analyze the feasibility of making interconnections of data spaces across different industrial sectors, using Catena-X as a reference, the main data space of the automotive sector.

This project will face different phases, including research on Catena-X's infrastructure as well as other data spaces, research on the possible interoperability methodologies to create communication channels among these data spaces, designing a solution to interconnect them, and implementing such design. Once the design is tested, it will be made more scalable, to ensure its applicability at a larger scale. Finally, the interconnectivity will be validated using a Digital Passport Product (DPP) model.

Chapter 2. TECHNOLOGY DESCRIPTION

One of the primary resources for this project will be creation and deployment of the Eclipse Dataspace Connector (EDC) configured for the Catena-X environment. This connector will enable a secure data exchange within the Catena-X ecosystem, serving as a testing ground for interconnectivity methodologies. Accessing the data space with the EDC will additionally require obtaining credentials for its federation within the Catena-X framework, a process supported by the Cofinity-X [1] platform. Cofinity-X can be seen as a 'Catena-X' spin-off, it is the open marketplace for Catena-X, bringing it to life. Cofinity-X is the platform which operates Catena-X. In other words, interaction within the Catena-X data space is carried out through its designated operator, Cofinity-X, which fully complies with all the standards defined by the Catena-X initiative.

To prepare for deployment and testing, several prerequisites will be addressed, including configuring environments with tools like Kubernetes, Docker, Maven, Terraform and JDK, and cloning the appropriate repositories from the EDC Tractus-X project [2]. These resources will validate be used to test connectivity and functionality, ensuring the feasibility of the proposed interconnection framework.

2.1 APPLICATIONS DEPLOYMENT TECHNOLOGIES

In order to deploy modern applications, there are several technologies that must be understood. The technologies explained in the following sections have been used throughout the project. In particular, in the deployment of the EDC, with the purpose of integrating Gestamp into Catena-X.

2.1.1 DISTRIBUTED PROCESSING AND CLUSTERS

Distributed systems are designed to solve computational problems that exceed the capacity of a single machine by leveraging multiple computers connected through a communication network. This model allows tasks to be divided and processed in parallel, significantly improving performance and scalability. However, distributed systems also introduce added complexity in terms of coordination, fault tolerance, and communication overhead. Nevertheless, they are still much more convenient than using a single machine to process data.

In a distributed system, the computing entities, called nodes or agents, can differ in roles and capabilities, and a group of these nodes working collaboratively forms a cluster. By distributing the workload, clusters enable higher throughput, improved fault tolerance, and better resource utilization. All this makes clusters and essential component in modern distributed architecture infrastructures.

2.1.2 CONTAINERS

In software engineering, containers are software units that package an application together with all its dependencies, libraries, and configuration files. This ensures that the application can run consistently across different computing environments. By encapsulating everything the application needs to function, containers eliminate common issues related to environment mismatches [3].

Containers implement a form of operating system-level virtualization. Unlike traditional virtual machines, containerized applications do not require a full guest operating system. Instead, they share the host operating system's kernel while remaining isolated from one another. This architectural approach significantly reduces overhead, since the host OS does not have to execute diverse operating systems at the same time. This is why containerization is also known as lightweight virtualization.

Moreover, containers encapsulate the application’s libraries and dependencies, keeping them isolated from those of other applications on the same host, which prevents version conflicts in an OS.

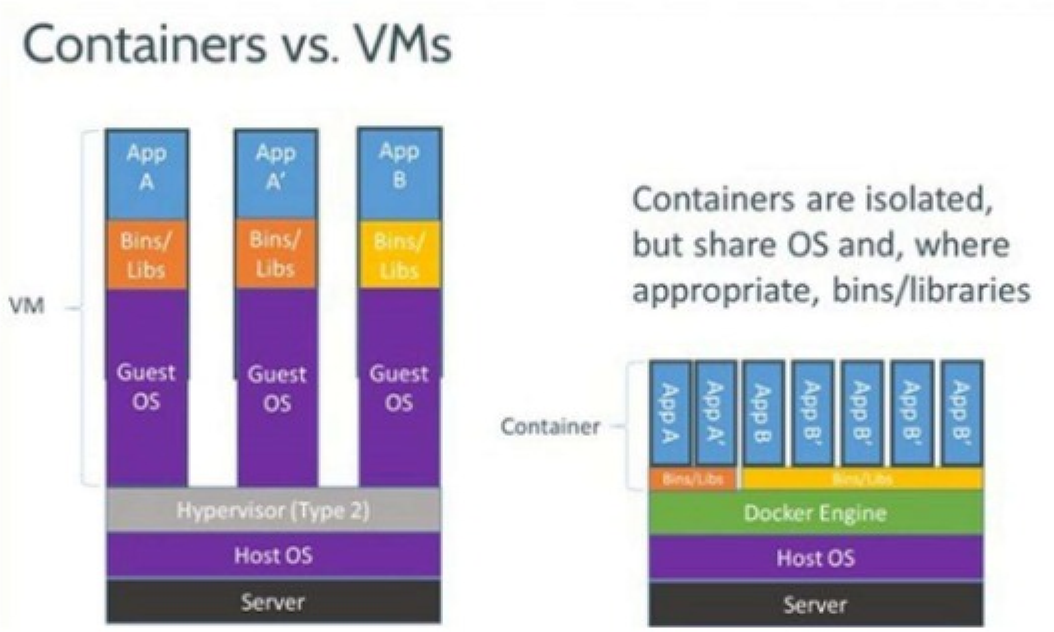


Figure 1. Containers vs VMs

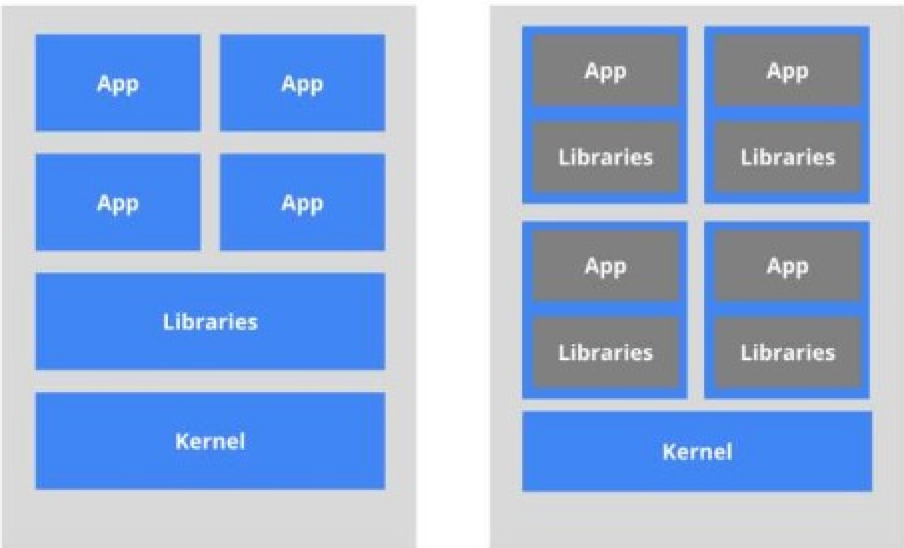


Figure 2. Containerization and encapsulation of apps & libraries

2.1.3 DOCKER

Docker is a platform which enables the creation, distribution and execution of containers. It offers tools to create container tools and execute them efficiently. Normally the containers run over the Docker engine, which is execute directly on the OS, as shown in Figure 1.

2.1.4 MICROSERVICES

Following the "divide and conquer" principle, software development becomes more manageable when applications are broken down into smaller, modular components. This is the foundation of microservices architecture, where each service focuses on a specific functionality and can be developed independently, while still collaborating with other services as part of a unique system [4].

Unlike monolithic applications, where all logic resides in a single block, microservices promote greater scalability. This architectural style aligns naturally with containerization: although not mandatory, it is considered a best practice to have one microservice per container, reinforcing modularity and ease of deployment.

2.1.5 PODS

Containers are grouped in pods. Often, different containers have very linked functionalities, so it makes sense to group them, since they will have to be executed at the same time. However, you can also have single container pods. In either case, containers are executed within pods, which are the minimal deployment unity in Kubernetes (explained later).

2.1.6 KUBERNETES

Kubernetes, also known as K8s, is an open-source platform that plays a key role in managing containerized applications. It acts as a container orchestrator, automating the deployment, scaling, and lifecycle of containers across a cluster of machines.

As mentioned before, Kubernetes groups containers into logical units called Pods and manages their execution across a cluster of computing nodes (worker machines). It handles scheduling, execution, and health monitoring. In the event of failures, Kubernetes can automatically restart containers or reschedule Pods, ensuring resilience and fault tolerance [5].

For tasks that are meant to run only once or need to complete a specific job—such as database migrations, backups, or batch processing—Kubernetes provides Jobs.

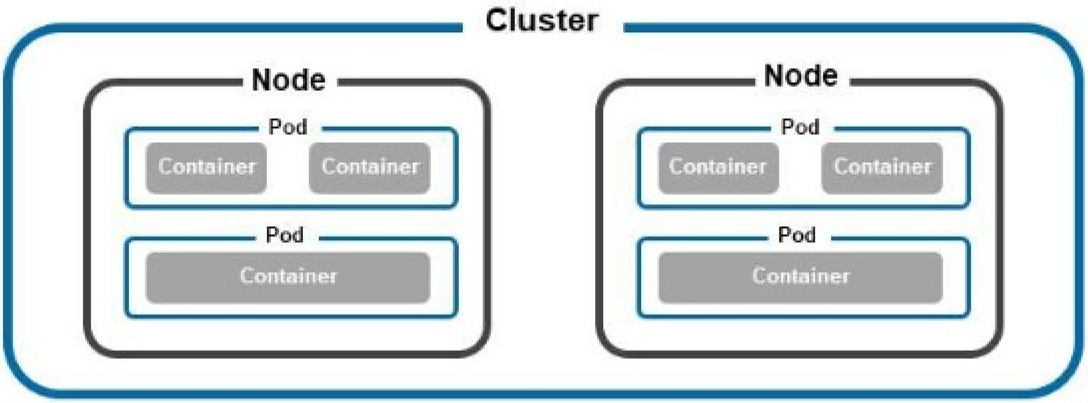


Figure 3. Components within a K8s cluster [6]

Figure 4 shows the lifecycle of containers, and the different technologies that have been emerging and evolving. First, applications were monolithic. The transition to microservices led to the need to develop containers. Therefore, Docker was born to execute them, and then Kubernetes was developed to orchestrate them.

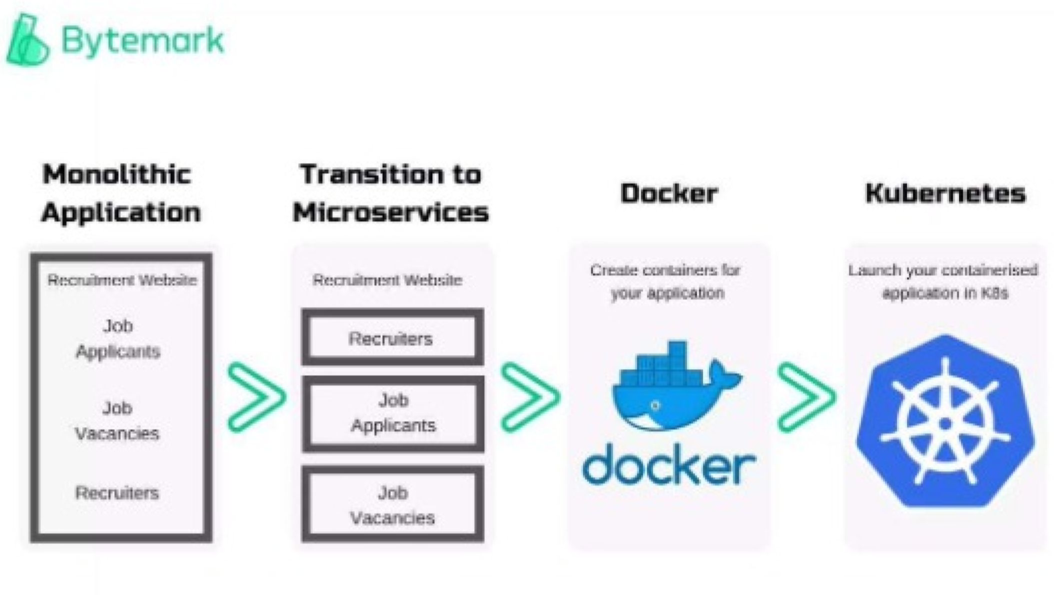


Figure 4. Containers lifecycle [7]

2.1.7 TERRAFORM

Terraform is another tool which is closely related to Kubernetes and containers. It is an IaC that provides tools to manage the resources (i.e. Kubernetes clusters) or the storage needed in a specific deployment. In other words, it is a resource manager for applications such as the one that will be deployed in this project (the EDC) [8].

Users can describe the infrastructure that will be needed in .tf files, written in HCL, and Terraform will provide with the resources.

Chapter 3. STATE OF THE ART

The exponential growth of data over the last decade has catalyzed the development of data spaces: decentralized platforms with the goal of securely enabling collaboration and data sharing across organizations while safeguarding data sovereignty. These federated spaces are crucial in fostering innovation, improving sustainability, and driving efficiency in complex industrial supply chains. Catena-X, a leading initiative in the automotive sector, exemplifies the transformative potential of data spaces [9].

Catena-X aims to establish a standardized, federated data ecosystem that enhances collaboration among manufacturers, suppliers, and other stakeholders in the automotive sector [10]. This decentralized architecture enables data sharing while allowing participants to retain full control over their data. By implementing uniform standards and protocols, Catena-X ensures interoperability and paves the way for a connected value chain that supports both operational efficiency and sustainability [11]. Its adherence to European data protection laws like GDPR underlines its commitment to data protection and security.

Catena-X promotes the development of technologies in the automotive sector such as digital twins, which allow real-time representation of physical assets and processes, enabling car manufacturers to simulate and optimize their products collaboratively [12]. Additionally, the use of the International Data Spaces Association (IDS) and Gaia-X (the European framework aimed at creating a federated and secure data infrastructure while upholding European principles of data sovereignty and privacy [13]), Catena-X secures data secure data sharing across its platform. It also aligns itself with the values of decentralization, transparency and trust [14].

However, the interconnection of data spaces across different industrial sectors (inter data space interoperability) has not yet been developed since it manifests significant challenges. The first is standardization; industries often use specific diverse data models and protocols, making inter-sector interoperability complex. The second challenge involves addressing regulatory differences, especially in data sovereignty and privacy requirements, which vary across industries and regions. Another key challenge is ensuring technical integration. Linking heterogeneous systems and architectures requires robust solutions capable of accommodating varying technological capabilities [15].

Catena-X is one of the most extended and successful data spaces among the Gaia-X framework. However, it is not the only example. Other known data spaces are EONA-X (Mobility, Transport and Tourism data space), Omega-X (energy sector data space), Factory-X (factories data space), or MDS (mobility data space).

Chapter 4. PROJECT VISION

4.1 MOTIVATION

The existing data spaces are extremely useful to enhance collaboration within their respective industries. However, this information remains sealed among each data space. In other words, there is no data exchange between different industries. This data isolation creates challenges for both suppliers and clients: suppliers face inefficiencies when attempting to evaluate cross-sector opportunities, as they must engage with multiple data spaces individually; and clients are restricted from gaining insights into the supply chains of other industries, limiting opportunities for innovation and optimization.

The ultimate motivation for this project is to create an interconnected ecosystem of data spaces across diverse sectors, transcending industry boundaries. Nevertheless, this interconnection presents the challenges previously described in the State of the art section. These challenges include the difficult interoperability and technical integration between data spaces, and the lack of trust among competing companies. However, collaborative efforts are being made, such as those by the International Data Spaces Association (IDSA), aiming to harmonize standards, promote trust, and create mechanisms for secure cross-sector data exchange. By developing gateways that enable a unified network of industrial data spaces, the project aims to overcome barriers to increase multi-sector data interoperability. Suppliers will gain a consolidated view of market dynamics and, at the same time, clients will be empowered to explore cross-industry synergies. As a consequence, innovation and efficiency in supply chains will be increased.

Furthermore, by fostering connectivity between data spaces, the proposed solution promotes the recycling of products across sectors, enhancing product sustainability. For instance, insights from the automotive industry could inspire innovations in other sectors, enabling the circular use of materials and reducing waste. Therefore, this project aligns with ESG global sustainability goals, supporting a more responsible use of resources in the supply chain of different sectors.

4.2 OBJECTIVES

The objectives of this project are structured into three categories to ensure alignment with its scope and deliverables. The high level general objective defines the overall purpose of the project. It is supported by intermediate level objectives, which represents specific goals that act as a bridge between the high-level vision and the tangible outcomes. Lastly, the low-level objectives represent those tangible outcomes, which are at the same time the concrete deliverables to be achieved.

4.2.1 HIGH LEVEL OBJECTIVE

The primary goal of this project is to design, implement, and validate a system that enables interconnectivity between data spaces from different industrial sectors. By leveraging Catena-X as a reference, this initiative seeks to enhance cross-sector collaboration, improve supply chain efficiency, and contribute to global sustainability efforts.

4.2.2 INTERMEDIATE LEVEL OBJECTIVES

There are some intermediate milestones that are set to achieve the high-level objective, which are the following:

- **Understanding current barriers**: Conduct an in-depth analysis of the Gaia-X infrastructure to understand the interoperability challenges in existing data ecosystems, including Catena-X and other industrial data spaces.

- **Developing interconnectivity solutions**: Designing and implementing a framework and methodology to facilitate data exchange across distinct data spaces.

- **Promoting sustainability**: Establish mechanisms that encourage the recycling and reusability of materials between industries, fostering innovation and reducing waste in line with ESG goals.
- **Testing scalability**: Ensure that the proposed interconnection solution can scale beyond initial prototypes to support broader industry adoption.

4.2.3 LOW LEVEL OBJECTIVES

These are the concrete, tangible goal deliverables to be met during the project:

1. Interconnecting three distinct prototype data spaces through a functional solution.
2. Enhancing the scalability of the implemented system to integrate additional data ecosystems beyond the initial three.
3. Validating the designed system using a Digital Product Passport (DPP). The DPP will serve as a component for ensuring the traceability of products, such as the Catena-X Battery pass [16]. By integrating the DPP into the validation process, the project will demonstrate how interconnected ecosystems can seamlessly share and leverage product data. This validation will not only test the feasibility of the proposed system but also highlight the benefits of cross-sector collaboration. The DPP enables a comprehensive analysis of how products from one sector could be recycled in another sector.

   Additionally, the DPP application aligns with global ESG goals by promoting transparency [17], optimizing resource use, and reducing waste through informed decision-making. The validation of the system using the DPP will reinforce its scalability and practical value.

4.3 METHODOLOGY & PLANIFICATION

The project will be divided into 4 different phases:

1. Data Spaces and Catena-X Research
2. Interconnection design
3. Interconnection implementation
4. Interconnection validation with DPP

Regarding the project planned schedule and methodologies, this can be found in Figure 5. As it can be seen, the project is planned to start on the 19^{th} November and finish on the 28^{th} April.

The development of this project has followed an Agile methodology. The Agile Manifesto [18] highlights the importance of developing adaptable and iterative working, having continuous feedback from stakeholders and frequent delivery of working components.

Instead of tackling the entire project at once, larger objectives (epics) are deconstructed into smaller user stories, which are then broken down into actionable tasks. These are executed within short, fixed-length iterations called sprints, three week long in this case.

Each sprint begins with a planning session to set clear objectives and assign responsibilities. At the sprint's conclusion, sprint review sessions have the goal of evaluating the output of the sprint, and retrospectives session to reflect on the process of the work.

Agile is grounded in Continuous Integration/Continuous Delivery (CI/CD), a practice that automates code testing and deployment to ensure that new features are released to users quickly and reliably.

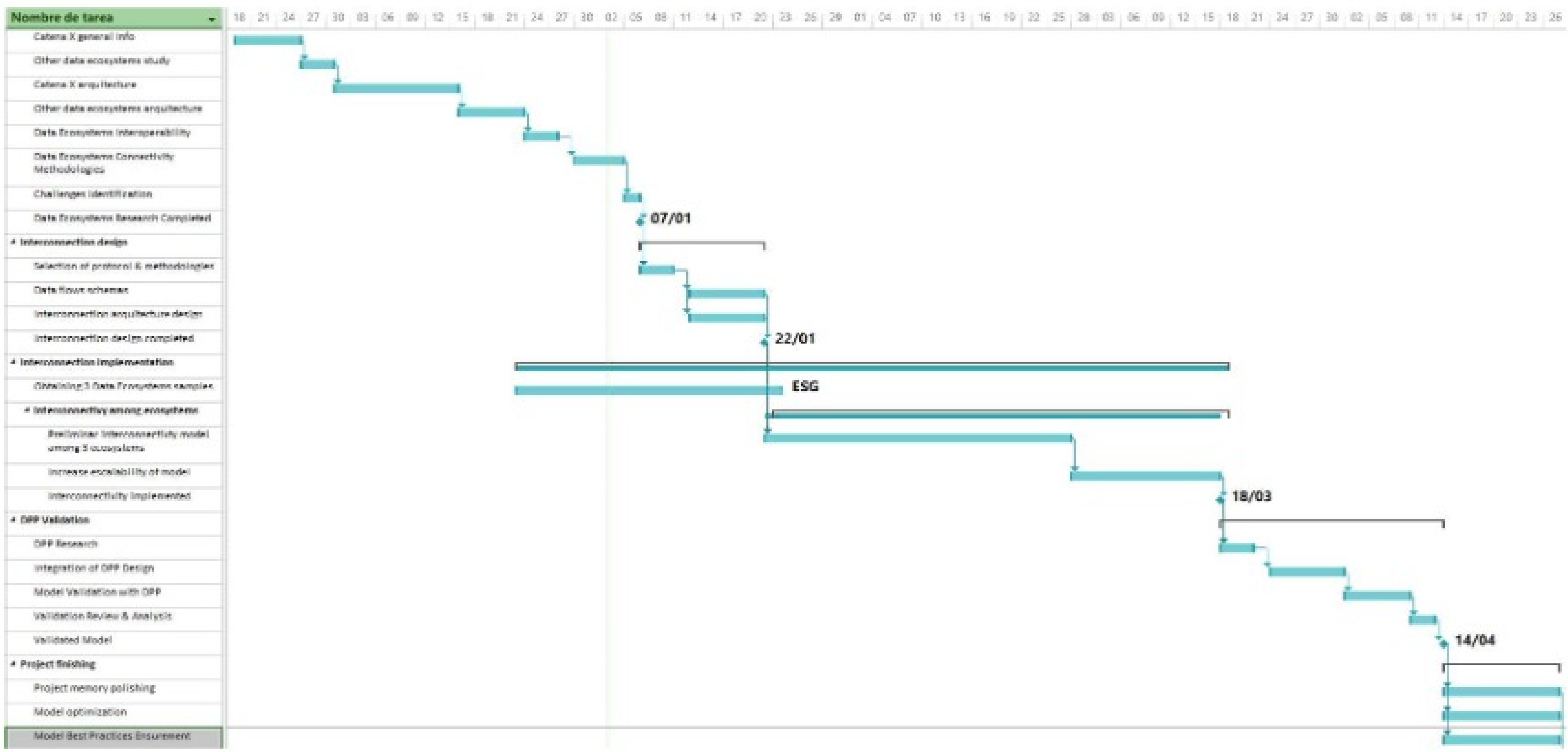


Figure 5. Project schedule

Chapter 5. DATA SPACES' OPERATING MODELS RESEARCH

To put the developed proposal in context, it is essential to first analyse the current operating models of the different dataspaces, and the existing data space frameworks. Once the engineering behind the data spaces is understood, the proposed interconnection will be explained.

Before the explanation is started, it is important to perform a brief introduction: Gaia-X is a conceptual framework which defines principles, norms, and architectures above which data spaces are built. It establishes a governance model for data spaces in Europe, specifying technical standards data spaces must comply with. Therefore, Gaia-X establishes the framework data spaces such as Catena-X or EONA-X later inherit and implement, adapting it to the needs of the industry in which they operate.

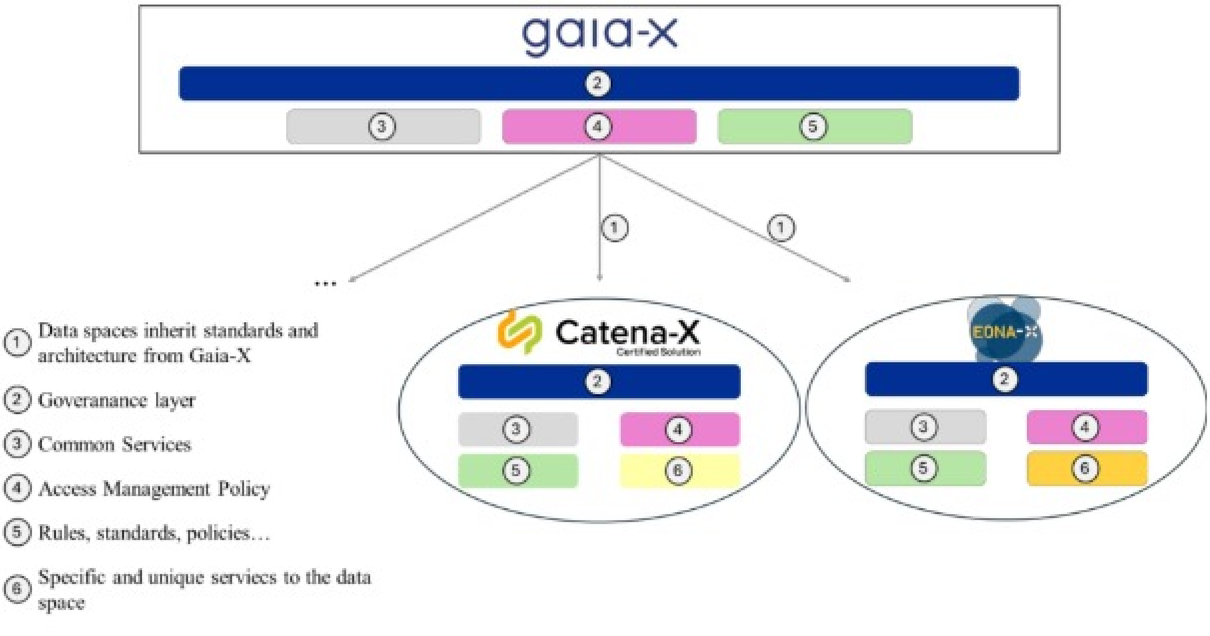


Figure 6. General overview of the data spaces environment

5.1 GAIA-X

Gaia-X is a European initiative designed to establish a secure and federated data infrastructure, enabling sovereignty, interoperability, and compliance with European regulations. It acts as a foundational framework for data spaces, such as Catena-X, and offers governance mechanisms to ensure trust and collaboration across the diverse stakeholders it has. Gaia-X emphasizes the core principles of openness, transparency, and security, aiming to unify disparate data models and structures under a single, coherent architecture.

However, Gaia-X is not alone in this initiative. Together with the International Data Spaces Association (IDSA), the FIWARE Foundation and the Big Data Value Association, they have founded the Data Spaces Business Alliance (DSBA) to "create value out of distributed digital (data) ecosystems".

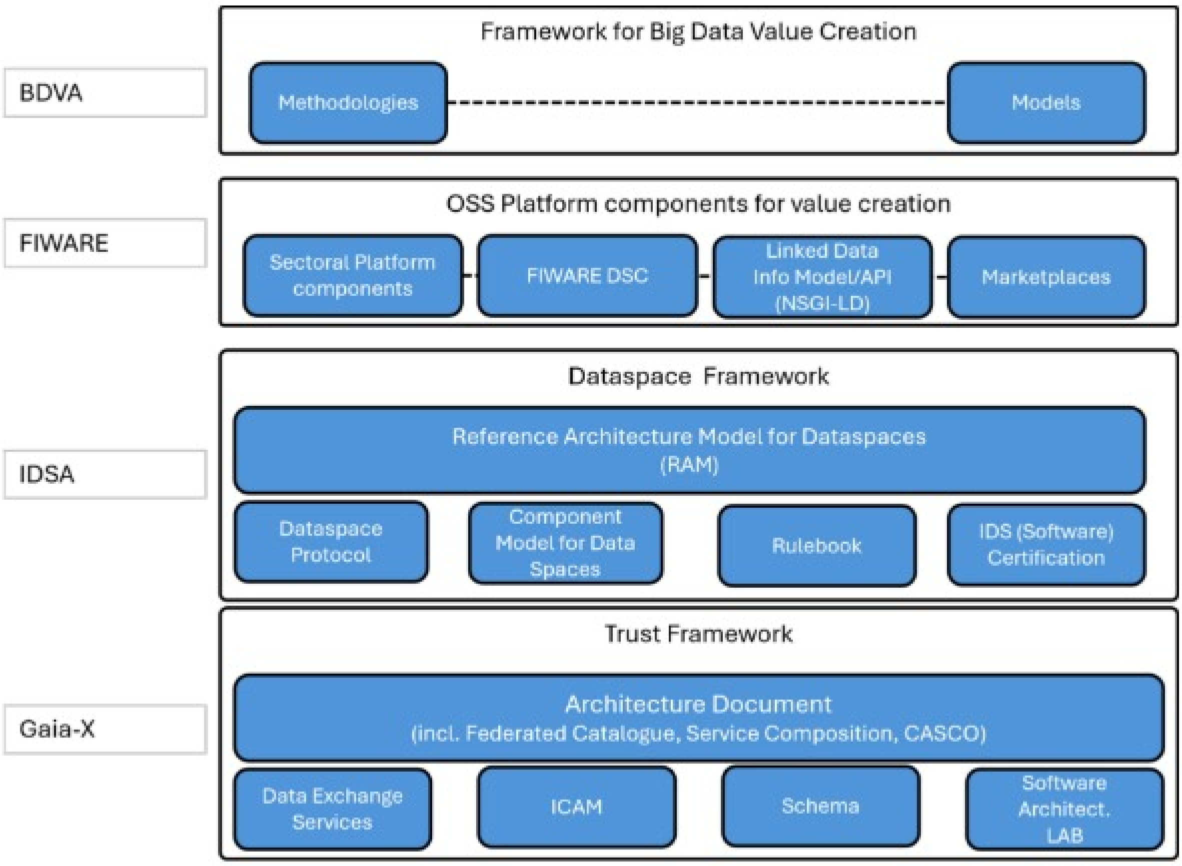


Figure 7. DSBA members [19]:

It is important to understand the key players behind Gaia-X. Although being a non-profit association under Belgian law, it has a mix of public and private stakeholders.

- Public stakeholders:
  - German government
  - French government
  - Other minor EU governments, such as the *Spanish Ministry for Digital Transformation and Civil Service.*

These public stakeholders have economic interests in the development and implementation of Gaia-X, which is also seen as a mechanism to keep the EU competitive in the global technological landscape.

- Private stakeholders (+300 companies):
  - Siemens
  - Microsoft
  - Bosch
  - Festo
  - SAP
  - Deutsche Telekom
  - BMW
  - …

Therefore, it can be stated that it has a mix of both public and private stakeholders.

Regarding the relationship between Gaia-X and the cloud, it is important to highlight the lack of European sovereignty in cloud infrastructure and the continent's heavy reliance on North American giant cloud providers. Equipping the EU with a sovereign cloud infrastructure will strengthen the independence and technological autonomy of European businesses, and Gaia-X can promote this change in paradigm.

GAIA-X is designed as an open and flexible data infrastructure. As such, all cloud providers are welcome to become stakeholders of the initiative at any moment, as long as they show commitment with the goals of GAIA-X and adhere to the UE rules for data sovereignty. GAIA-X maintains a neutral stance regarding cloud providers; it does not favor any particular platform, not even major players like AWS, Azure, or Google Cloud. Instead, it offers equal opportunities for both global and smaller European providers to participate. This inclusive approach enables collaboration with European companies such as IONOS, SAP S/4HANA Cloud, and Arsys, among others.

This is why GAIA-X provides strong support for European providers: it empowers them to better compete with global cloud American giants. This contributes to a more competitive cloud market and strengthens the European technological independence.

5.1.1 GAIA-X | THE OPERATING MODEL

Gaia-X's operating model is structured around governance, compliance, and roles, all interconnected with each other. At its core is the Gaia-X European Association for Data and Cloud AISBL, headquartered in Brussels, which oversees the initiative's policy, compliance, and technical implementation. The AISBL develops the technical framework and operates the Gaia-X Federation Services. [20]

5.1.1.1 Roles and governance

The Gaia-X ecosystem includes roles categorized into service providers, consumers, and operators, each contributing to the secure exchange of data. Participants operate under a governance model that emphasizes decentralization and trust, supported by Trust Anchors. Trust Anchors provide the foundational mechanisms and policies to establish a trusted ecosystem but do not directly certify participants or services. Instead, certification and compliance verification are carried out by independent Conformity Assessment Bodies (CABs), which ensure participants and services adhere to the Gaia-X framework. [21].

The AISBL also includes several Gaia-X Committees, which define policies, technical standards, and compliance requirements. These committees include the Policy Rules Committee (PRC), Technical Committee (TC), and Business Committee, which collaboratively ensure the alignment of Gaia-X standards with European values [22].

5.1.1.2 Compliance and labelling

Gaia-X employs a three-level compliance system depending on how much an entity adheres to the adherence to the Gaia-X standards:

"Gaia-X distinguishes 3 levels of Labels, starting from Label Level 1 (the lowest), up to Label Level 3 (the highest), which represent different degrees of compliance with regard to the goals of transparency, autonomy, data protection, security, interoperability, portability, sustainability, and European Control" [22].

It is important to note that the labelling system is modular, allowing new labels to address evolving needs, such as sector-specific or geographic requirements.

5.1.2 GAIA-X | THE ARCHITECTURE

5.1.2.1 Conceptual layers

Gaia-X's architecture is organized into three layers, each addressing specific requirements [19]:

1. Ecosystem governance: A framework of rules established by the AISBL and agreed upon by all participants in the ecosystem, which must be implemented. This governance defines a group of providers responsible for delivering foundational ecosystem services, known as Federation Services.
2. Infrastructure: The physical and digital resources, including computing, storage, and network services, that adhere to the governance rules offered by the service providers.
3. Data Ecosystems and Data Spaces: Where participants comply with the governance rules, leveraging the infrastructure to access and share data. For example, Catena-X.

To summarize, Gaia-X's architecture is structured into the Ecosystem Governance, a rule framework implemented by participants, the Infrastructure with the physical and digital resources provided following the Ecosystem Governance rules, and the Data Ecosystems and Spaces that use the infrastructure.

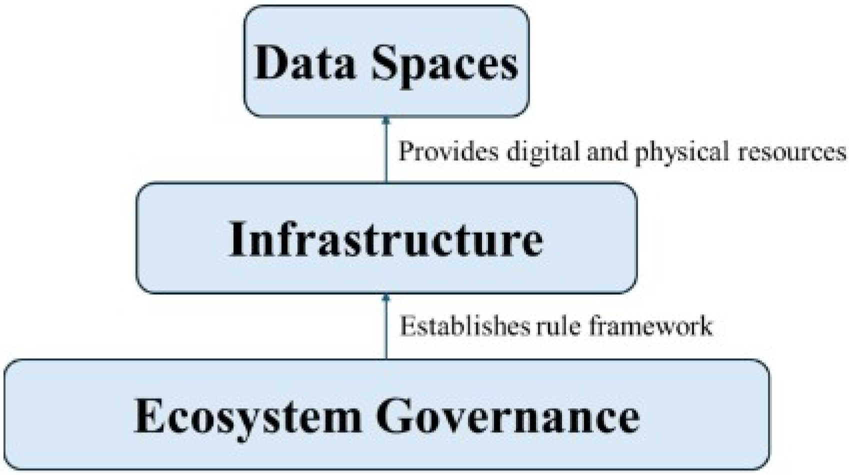


Figure 8. Gaia-X architecture conceptual layers

5.1.2.2 Interactions of Participants

Once the theoretical architecture is understood, it is important to analyse how do participants interact with each other within the data space, making use of the Infrastructure and following the Ecosystem Governance.

Firstly, both the data providers and the consumers must have their credentials and be validated to interact with each other.

5.1.2.2.1 Validation and certification

Gaia-X employs a rigorous validation process for participants and services [23]:

- Application: Participants submit their compliance documentation to receive the needed credentials to participate in the ecosystem. In order for participants to be candidates, they first have to own a Decentralized Identifier (DID) to be uniquely identified.
- Assessment: Independent CABs evaluate the submissions and verify if the applications meet the policy rules set by Gaia-X.
- Certification: Once validated, participants receive Verifiable Credentials (VCs). DIDs which include VCs are known as Trusted IDs. These Trusted IDs act as a digital passport, enabling participation in the ecosystem. Trusted IDs are

presented to the Gaia-X Digital Clearing House (GXDCH). Once verified by the GXDCH, the participants are allowed into the ecosystem.

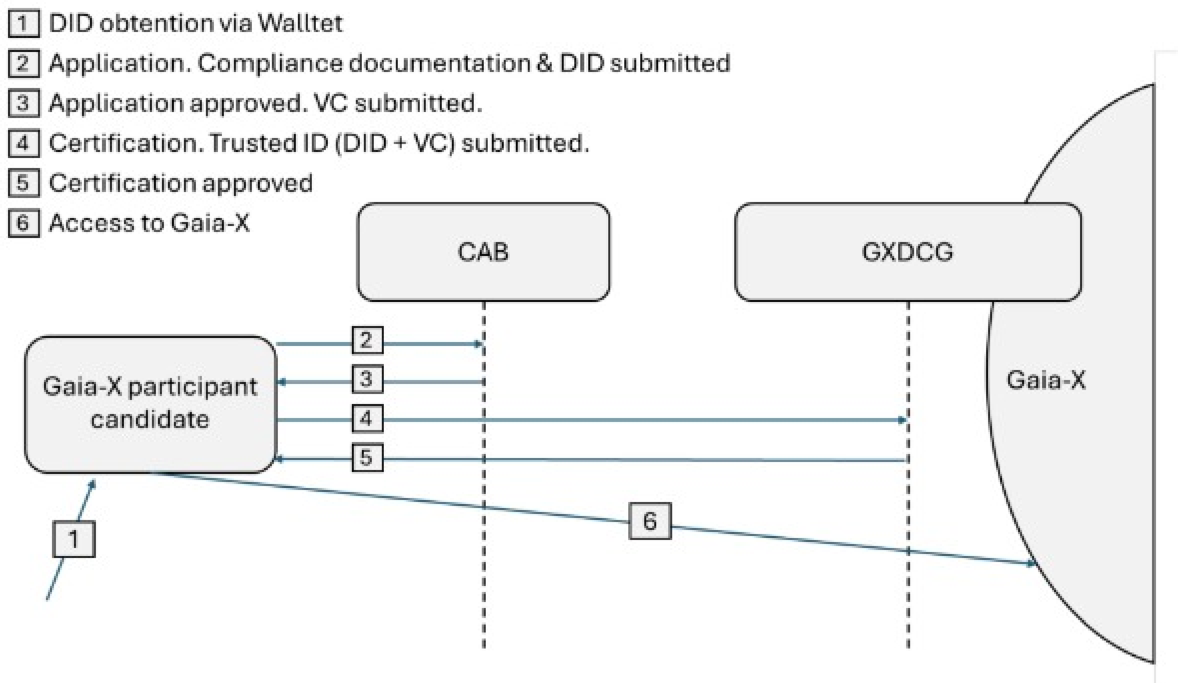


Figure 9. Gaia-X validation and certification process

This validation and certification mechanism is foundational to establishing trust in the ecosystem and is known as the ecosystem's Identity, Credential and Access Management (ICAM) system. It ensures that only authorized participants access the ecosystem and that all interactions adhere to compliance [23].

Gaia-X integrates OpenID Connect (OIDC) [24] for federated identity management. OIDC is an authentication protocol based on the OAuth 2.0 standard that allows users to securely authenticate across different applications and services using a DID. This integration allows participants to use a single digital identity across the ecosystem, streamlining authentication processes and enhancing user experience.

5.1.2.2.2 Gaia-X services

The Data Consumer begin with the creation of a declaration, where they provide their credentials. Once the declaration is made, the Consumer must maintain their Gaia-X credentials, ensuring that they remain valid and up-to-date. Similarly, Providers also begin by creating a declaration, which involves detailing their credentials, the services they offer, and their compliance with the ecosystem's standards. By doing so, they ensure that their Trusted ID remains reliable and builds confidence with potential Consumers.

The GXDCH acts as a central repository for validated participants and their credentials. If the GXDCH verifies the VC in the Trusted ID of the participants, the participants are certified into the Data Space, and their information is stored in the Gaia-X Registry within the GXDCH. Once consumers and providers are certified, providers describe and list their offerings in a Gaia-X Catalogue (which acts like a marketplace). Consumers search for the catalogues offerings that match their needs. Once a match is found, the Consumer and Provider negotiate a contract. It is important to note Gaia-X does not interfere in the negotiation but ensures that both parties are trustworthy through the GXDCH VC's verification.

The Gaia-X European Association for Data and Cloud AISBL plays a foundational role in the ecosystem by defining the rules and standards that govern participant interactions. It begins by establishing the technical and interoperability requirements for services and participants. The Association also defines policies that guide the governance and operational aspects of the ecosystem. These policies ensure that all interactions remain transparent, secure, and compliant.

It is important to understand the difference between CABs, the GXDCH and the Gaia-X AISBL.

Aspect	CAB	GXDCH	Gaia-X AISBL

Role	Certification and compliance assessment.	Credential repository, verification and storage	Ecosystem governance and overseight.
Independence	Operate independently to Gaia-X but under AISBL rules.	Acts as a centralized service within the ecosystem.	Governs the entire ecosystem.
Decision-making	Decide if a participant complies with a standard.	Verifies the VC is valid and stores participant's information.	Defines standards, rules and strategies.
Credential Issuance	Yes, issue VC	No, verifies VCs	No

Table 1. Differences between CABs, GXDCH and the Gaia-X AISBL

Finally, ecosystem operators are responsible for ensuring compliance within particular ecosystems. They oversee the operation of ecosystem compliance, verifying that the services offered by Providers and used by Consumers adhere to the agreed-upon standards. In addition, they handle the issuance of ecosystem-specific credentials, which are tailored for particular agreements or data-sharing contexts and use cases.

5.1.2.3 Data Exchange & Data Products

The participants interact by sharing data. Data exchange in Gaia-X refers to the structured and secure sharing of data between participants (e.g., data providers, consumers, and intermediaries) within the ecosystem. The focus is not just on sharing data but doing so in a way that ensures trust and compliance with the Gaia-X standards. In Gaia-X, data is exchanged through Data Products [19].

Each Data Product in Gaia-X represents a package of data coming from the providers towards the consumer that has been prepared, enriched, and made available for use within the ecosystem. Each Data Product is standardized compliant with Gaia-X terms, and contains:

- Metadata that describes its purpose, structure and how it can be used
- Specific governance rules defined in contracts and licenses about specific needs
- Traceability information

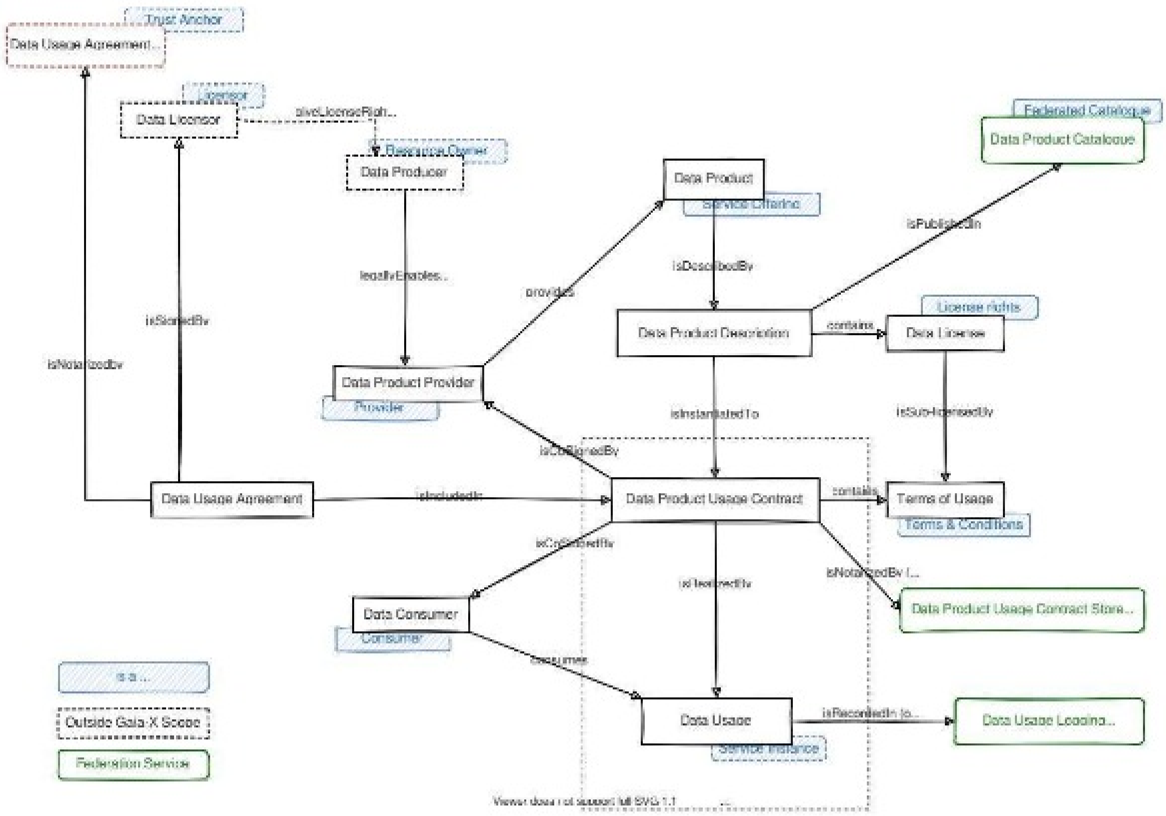


Figure 10. Data Product conceptual model [19]

Figure 10 illustrates the structured flow and interactions involved in the lifecycle of a Data Product within the Gaia-X ecosystem [19].

At the foundation of the model is the Data Producer, the entity responsible for generating raw data. This data could originate from various sources, including sensors, enterprise systems, or manual input, depending on the specific use case.

The next participant in the flow is the Data Provider, which plays a pivotal role in transforming the raw or enriched data into a Data Product. The provider ensures that the data is curated and supplemented with the necessary metadata, governance rules, and technical specifications required to make it interoperable within the Gaia-X framework. Once the Data Product is prepared, the provider publishes it in a Gaia-X Catalogue, a federated and standardized marketplace for data and services. The publication process includes making the Data Product discoverable by potential consumers, adhering to established governance rules, and ensuring that it complies with ecosystem standards. The Data Provider essentially bridges the gap between raw data generation and its practical application by making data usable and accessible within the ecosystem.

The Data Consumer interacts with the Gaia-X Catalogue to discover Data Products that align with their needs. Once a suitable Data Product is identified, the consumer engages with the provider to establish a contractual agreement.

A critical step in the flow is the negotiation and establishment of the Data Product Usage Contract. This agreement forms the legal framework governing the Data Product's use. They specify the conditions under which the data can be accessed and utilized, such as time limitations, specific use cases, and restrictions on sharing or modifying the data. Only after the contractual obligations are met, the Data Consumer accesses the Data Product using secure mechanisms, such as APIs or dedicated data platforms.

5.1.3 ROLE OF GAIA-X IN DATA SPACES

Gaia-X serves as a unifying framework for data spaces, providing the foundational architecture, governance, and compliance mechanisms. Data spaces like Catena-X leverage their architecture based Gaia-X to ensure their operations align with European values of data sovereignty, security, and interoperability.

For example, Catena-X illustrates how Gaia-X supports sector-specific ecosystems. By adhering to Gaia-X standards, Catena-X ensures, in the first place, interoperability, with standardized APIs which enable data sharing across the ecosystem's participants. In the second place, it ensures compliance, since the participant's credentials are verified by the GXDCH. In the third place, there is data sovereignty, where participants retain full control over their data, with robust mechanisms for managing access and usage policies.

In conclusion, Gaia-X provides the structural and operational foundation for data spaces. It determines the infrastructure used by data spaces in different industries, such as Catena-X in the automotive sector.

A valid object orientated programming analogy to understand the relationship between Gaia-X and data spaces such as Catena-X or EONA-X is to see Gaia-X like a programming class which defines a set of rules, properties and behaviors that the instances must inherit. Gaia-X provides a framework with governance rules and specifications. It is abstract and generic, and it can be applied in various sectors and ecosystems. Catena-X (or EONA-X) is an "instance" of Gaia-X which inherits the behaviors Gaia-X has applying them to the automotive sector, adding particular features such as the Use Case KITs which will be analyzed Catena-X | Use Cases and KITs.

5.2 CATENA-X

Catena-X is the most implemented data space within the Gaia-X infrastructure, and it implements the framework developed by Gaia-X. As one of the leading data space project, its structure will be analyzed and studied to understand the key elements that define a successful data space.

5.2.1 CATENA-X | THE OVERALL ENVIRONMENT

Catena-X environment consists of three environments:

Catena-X Automotive Network e.V., the development environment, and the operating environment.

1. CatenaX Automotive Network e.V. ("the Association"): In charge of standardization, coordination and management of the data space. Members can participate in

committees and shape the Catena-X ecosystem. It publishes standards to ensure interoperability, which must be complied by the members to ensure their participation in the data space. Catena-X standards, however, are based on Gaia-X AISBL and in the International Data Space Association (IDSA).

2. Development environment: Research is made to develop candidate standards in the Development environment, which are then approved by *the Association.*

3. Operating environment: Where the application of the services developed by the Development environment takes place. Different services are operated by different provides.

5.2.2 CATENA-X | THE ARCHITECTURE

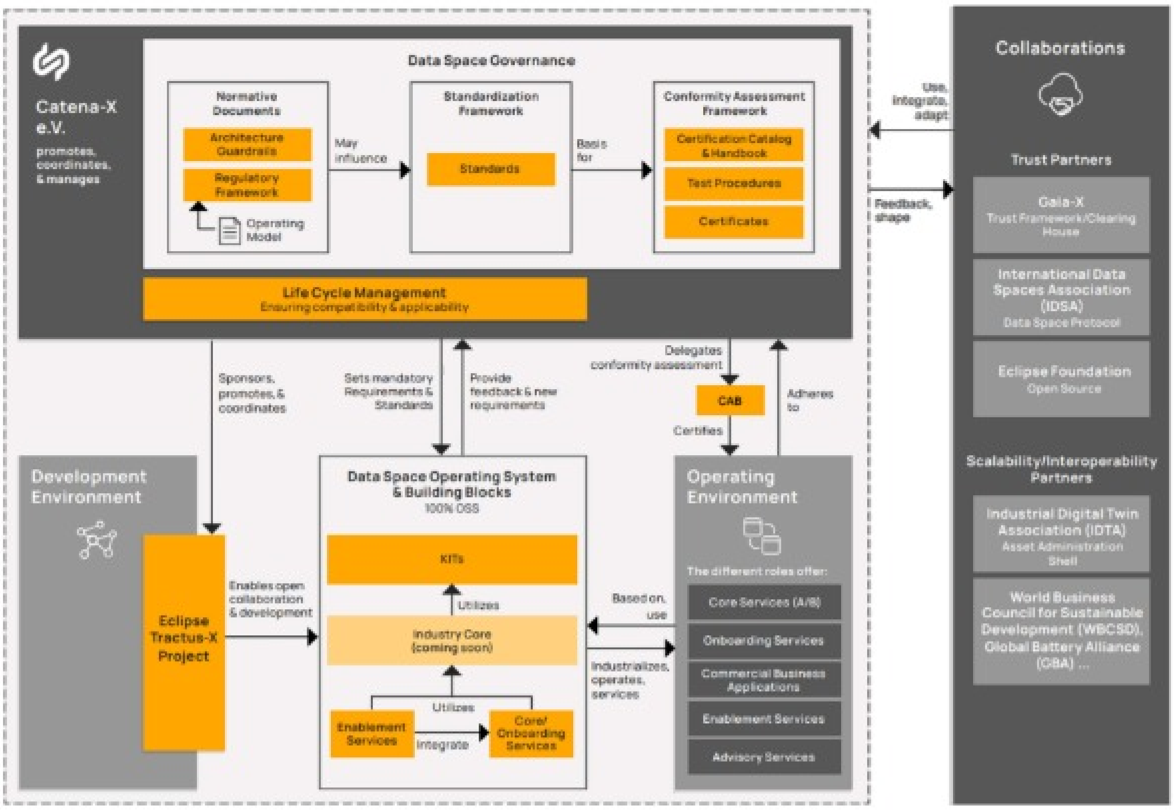


Figure 11. The Catena-X architecture [25]

5.2.2.1 The Association as the 'Government' of Catena-X

As previously explained, the Catena-X Association serves as the main management body of the Catena-X ecosystem. Similar to the Gaia-X AISBL, it operates as a governance entity. Its responsibilities include: coordinating and promoting particular standards for the automotive sector (via the Standardization Framework); managing regulatory processes through documents like the Regulatory Framework and Architecture Guardrails; and overseeing interoperability and ensuring that all participants operate under a shared framework.

5.2.2.2 The Association Promotes Eclipse Tractus-X

The Association promotes technical development and standardization through the Eclipse Tractus-X project. This project operates within the Development Environment, researching and developing new possible standards, and creating open-source solutions through the Building Blocks. These blocks are developed by the same Eclipse Tractus-X project and form the foundation of services and solutions within Catena-X.

To summarize, the Eclipse Tractus-X operates in the Development Environment. It creates solutions to improve Catena-X, known as Building Blocks. When approved, these solutions are implemented in the Operating Environment.

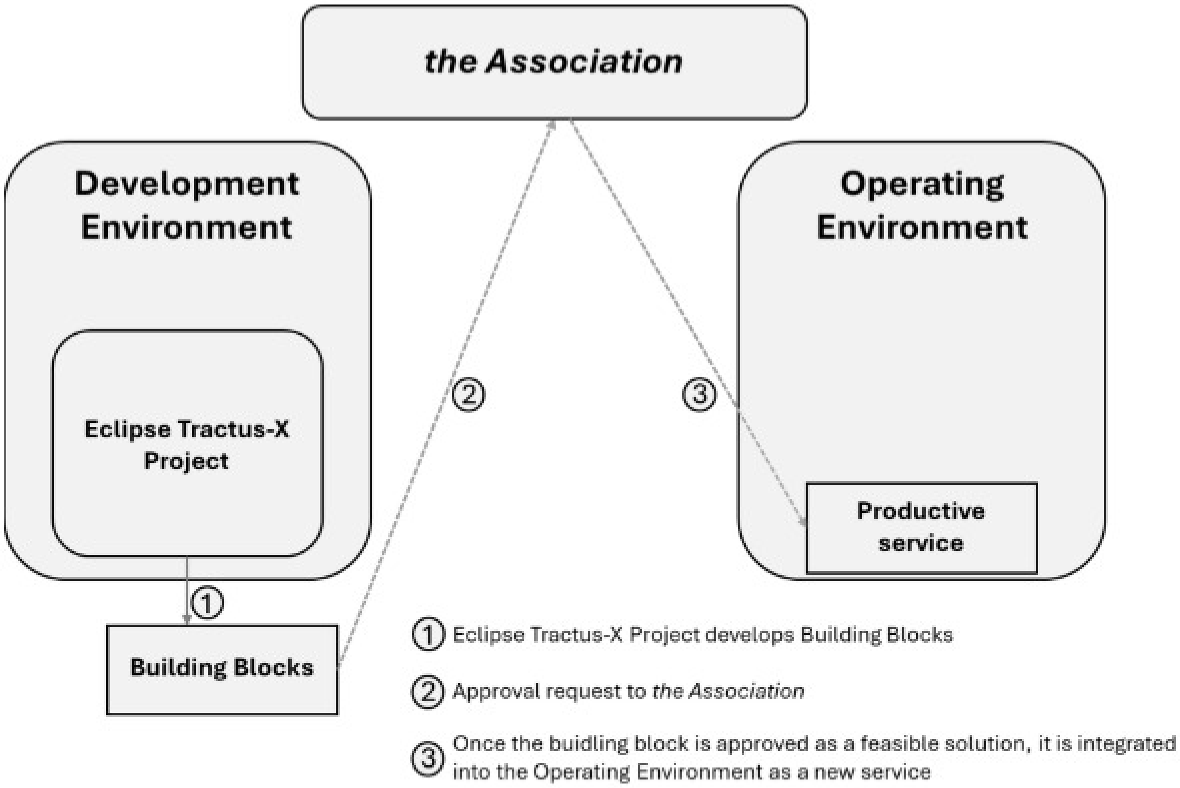


Figure 12. From building blocks to Operating services

5.2.2.3 Building blocks

The main Building Blocks are the following:

- Enablement Services and Onboarding Services:
  1. Enablement Services: Provide advanced tools and services to facilitate technical integration, such as tools for data interoperability, digital twins management or the EDC [26].
  2. Onboarding Services: Help companies join the ecosystem by ensuring compliance with technical and regulatory requirements. They include registration, validation, and initial technical support.
- Industry Core:

  Represents the specific needs of the automotive industry. Still in development.
- Use Case KITs:

These are toolkits designed for specific use cases. They include standards, software components, and guides to implement interoperable solutions in the ecosystem. These KITs will be described in Catena-X | Use Cases and KITs.

5.2.2.4 The Operating Environment and Marketplaces

In the Operating Environment, the Building Blocks developed in the Development Environment are implemented as services and applications. These services are offered in certified marketplaces. Meanwhile, *the Association* delegates independent CABs the task of issuing certificates to the appropriate services offered at the marketplaces.

5.2.2.5 Services in the Operating Environment

The Operating Environment creates various services, some of which are the Core Services A/B which rovide foundational services like identity management, participant discovery, and data registration; Onboarding Services; Enablement Services; Advisory Services which offer consulting to help participants maximize the benefits of the ecosystem; and Commercial Business Applications which are industry-specific applications developed by third parties to address particular use cases like traceability, sustainability, and supply chain optimization.

More will be described in Catena-X | Role Management.

5.2.2.6 Collaborations

The trust partners include Gaia-X, the International Data Spaces Association (IDSA), and the Eclipse Foundation.

Gaia-X provides the trust framework for federated and interoperable data spaces, promotes compliance with European data sovereignty principles (and rules), and includes the GXDCH [27] which must verify all the onboarding candidates.

The IDSA supplies the architecture principles that enable data exchange between participants and defines standards and protocols for secure and interoperable data transactions.

The Eclipse Foundation hosts the Eclipse Tractus-X project and manages its open-source development. However, considering it merely a collaborator may be ambiguous—it is much more than that. It is a key component in the development of solutions within the Development Environment. Without Eclipse Tractus-X, for example, the Catena-X EDC as we know it would not exist, and the EDC is a fundamental element of the ecosystem, as it will be seen later.

Regarding scalability and interoperability partners, the ones that can be found are the Industrial Digital Twin Association (IDTA), which provides standards to enable digital twin management; and the World Business Council for Sustainable Development (WBCSD) and Global Battery Alliance (GBA), which provide scalable data models for cross-industry use cases.

5.2.2.7 Value created

Such architecture ensures that Catena-X provides the participants with an ecosystem of the following characteristics:

- Interoperable: Ensures data exchange between different stakeholders in the automotive value chain.
- Data Sovereign: Maintains participants' control over their data.
- Optimized: Common standards and solutions improve efficiency in supply chain operations.
- Innovative: Use Case KITs foster the development of new solutions and business opportunities.
- Trust: Governance frameworks and certified marketplaces establish a trusted environment for data exchange.

5.2.3 CATENA-X | USE CASES AND KITS

Regarding the Catena-X Operating Model description, *"The goal of a Catena-X use case is to solve a specific business problem and to create value for data providers and consumers"* [25]. Catena-X uses KITs for this purpose.

KITs (Keep It Together) bundle all the requirements and resources needed by the ecosystem's participants for a specific use case. All KITs, independently to the use case to which they are orientated, are always structured in the same way to ensure standardization, and include the following:

1. Semantic models for data understanding
2. APIs for technical interoperability.
3. Logic/schemas for business calculations.
4. Deployment scripts and guides for easy integration.

For example, if a company adheres to Catena-X to exchange Product Carbon Footprint (PCF) data, they will use the standardized Carbon Footprint KIT to obtain the appropriate software, components, standards and accesses to exchange carbon footprint data with other use case participants.

5.2.4 CATENA-X | ROLE MANAGEMENT

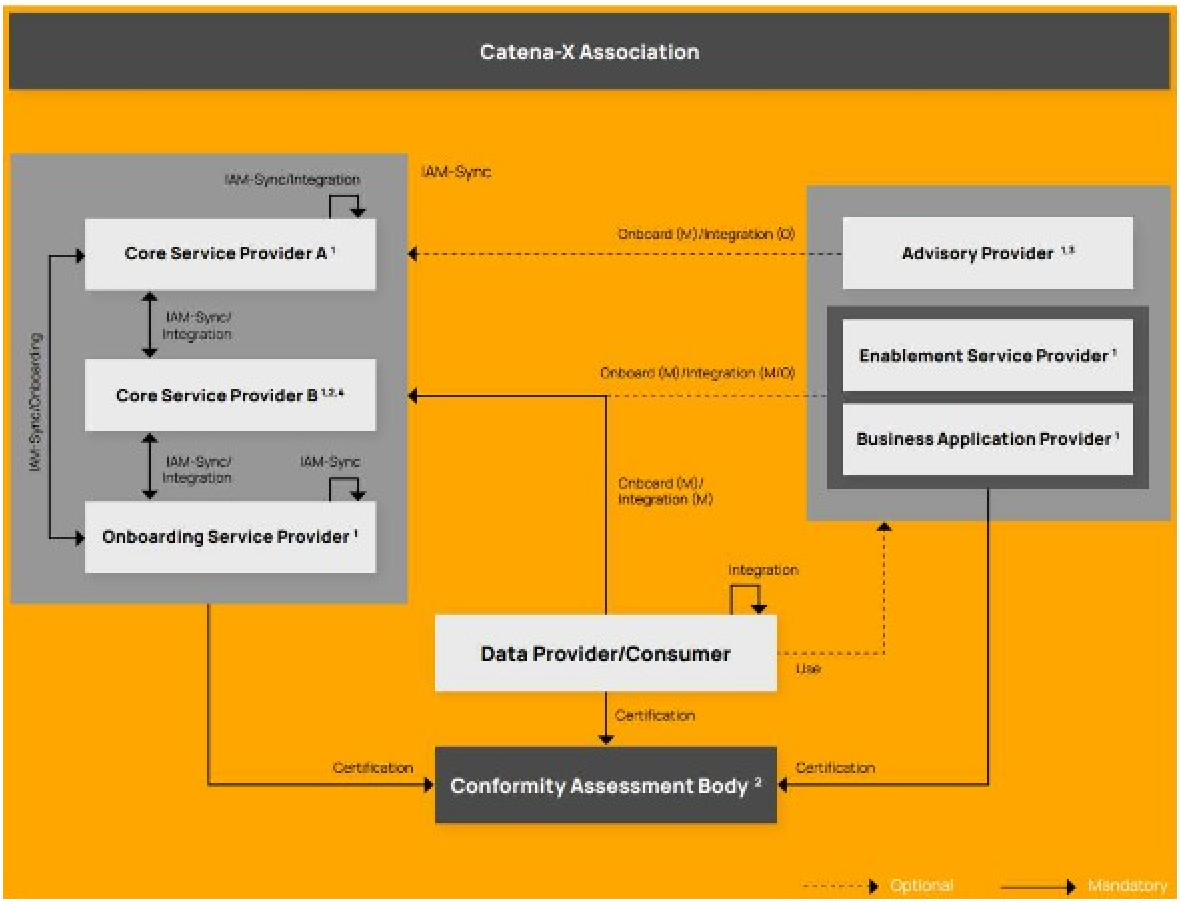


Figure 13. Catena-X Roles [20]

As previously mentioned, there are different roles and services in the Operating Environment. Roles use KITs to implement solutions aligned with specific use cases.

5.2.4.1 The Roles

Such architecture ensures that Catena-X provides the participants with an ecosystem of the following characteristics:

- CSP-A: Operates and offers services like marketplaces, semantic hubs, and business partner information systems. KITs solutions are hosted in the CSP-A marketplace.

- IAM Integration [1]: Must synchronize IAM among CSPs to maintain secure identity and access controls.
- Relationships: Integrates with CSP-B, Onboarding Service Providers(OSPs,) and other CSP-As.
- Prerequisites: Certified by Conformity Assessment Bodies (CAB) and compliance with Catena-X standards.

- CSP-B: The central role in the ecosystem, managing critical services such as identity wallets and discoverability functions. It is unique, ensuring consistency across the ecosystem.
    - IAM Integration: Acts as the central IAM system, connecting all participants securely.
    - Relationships: Integrates with CSP-As and Onboarding Service Providers (OSPs), and other roles.
    - Prerequisites: Nominated by the Catena-X Association, certified by CAB, and compliant with all frameworks.

- OSP: Helps new participants register and integrate technically into the data space.
    - IAM Integration: Links participants to the central IAM for identity validation.
    - Relationships: Collaborates with CSP-B and Gaia-X Clearing House for participant validation.
    - Prerequisites: Certified by CAB and must use the Gaia-X Clearing House.

- Enablement Service Providers (ESP): Provides tools for seamless data exchange, mapping the providers'/consumers' data to the Catena-X semantic models.

[1] Note that Catena-X uses IAM (Identity and Access Management), and Gaia-X uses ICAM (Identity, Credentials and Access Management). This is because Gaia-X manages all the ecosystems' credentials through the GXDCH, so the particular ecosystems are not in charge of that task.

Therefore, it is the service in charge of creating the Data Product (i.e acts as a Data Provider) mentioned in Data Exchange & Data Products.

- IAM Integration: Uses CSP-B's identity services for secure operations.
- Relationships: Works with CSP-As, CSP-B, and OSPs.
- Prerequisites: Certification by CAB.

- Business Application Providers (BAP): Offers applications for specific use cases like carbon footprint analysis, making use of KITs.
  - IAM Integration: Relies on CSP-B for secure data access and exchange.
  - Relationships: Collaborates with CSP-As and uses CSP-B and OSP services.
  - Prerequisites: Must be certified, compliant with standards, and listed in a CSP-A marketplace.

- Advisory Providers (AP): Provides strategic and technical advice to optimize participation in Catena-X.
  - Relationships: Uses services from CSP-As and CSP-B where needed.
  - Prerequisites: Qualified by the Catena-X Association.

- Data Providers and Consumers (DPC): Share and process data to address industry challenges.
  - IAM Integration: Must use CSP-B's IAM for secure and compliant data exchange.
  - Relationships: Collaborates with CSPs, ESPs, BAPs, and APs.
  - Prerequisites: Uses certified services and complies with the framework.

- Conformity Assessment Bodies (CAB): Certifies participants and services to validate their alignment with Gaia-X (and therefore, Catena-X) standards.

It is worth mentioning the relationships between use cases, KITs, and roles. Use Cases define the business problem and value proposition (e.g., sustainability); KITs standardizes the solution with tools, schemas, and processes for interoperability; and Roles implement and deliver the KIT through services in the Operating Environment: CSP-A hosts KIT-based solutions in marketplaces, BAP develop applications using KITs, ESP provides infrastructure for KIT implementation and CAB certify KIT components and solutions.

All these processes and roles' management are governed by the Association.

Role	Relationships	Prerequisites
CSP-A	CSP-B, other CSP-A, OSP	Certified by CAB; complies with standards.
CSP-B	All other roles	Nominated; certified by CAB; compliant.
OSP	CSP-B, Gaia-X Clearing House	Certified; integrates Gaia-X services.
ESP	CSP-A, CSP-B, OSP	Certified by CAB.
BAP	CSP-A, CSP-B, OSP	Certified; listed in CSP-A marketplace.
AP	CSP-A, CSP-B	Qualified by the Association.
DPC	CSPs, ESPs, BAPs, APs	Certified services and applications.

Table 2. Role relationships and prerequisites summary

5.2.4.2 Example: Battery Manufacturing and Data Collaboration

5.2.4.2.1 Scenario

A battery manufacturer wants to enhance the sustainability of its supply chain by collaborating with a supplier to reduce the carbon footprint of materials. They need to share and access data securely while ensuring compliance with regulatory and technical standards.

5.2.4.2.2 Step by step using Catena-X

1. Joining the ecosystem:

   The manufacturer must integrate into the Catena-X ecosystem through the Onboarding Process.

   a. The manufacturer collaborates with an Onboarding Service Provider (OSP) to register their organization. The OSP then validates their credentials using the GXDCH, ensuring they comply with identity and regulatory standards.
   b. Once validated, a Business Partner Number (BPN) and a digital identity wallet are issued by CSP-B, enabling secure and trustworthy interactions within the ecosystem.

2. Integrating into the ecosystem:

   The second step involves Technical Integration.

   a. The manufacturer deploys the Enablement Services, which include an EDC to facilitate standardized data exchange and the Identity Wallet provided by CSP-B, which is used to manage authentication and control access to data.

3. Defining the Use Case: Sustainability:
   a. The manufacturer selects the PCF Use Case, which focuses on calculating and reducing the environmental impact of battery materials.
   b. To support this use case, a specific KIT is provided that includes several components: semantic models that define emissions data, APIs that enable standardized communication between participants, logic and schemas for calculating the PCF, and deployment scripts that assist in integrating these components into the manufacturer's systems.
   c. This KIT is available through the Catena-X certified marketplace, which is managed by CSP-A.

4. Data Exchange with the Supplier:
    a. The supplier also integrates into the Catena-X ecosystem through an OSP, ensuring compliance with network requirements and obtaining a digital identity wallet issued by CSP-B.
    b. Once both parties are onboarded, the data sharing setup begins. The manufacturer and supplier establish a secure connection using their respective EDCs. Leveraging the KIT's semantic models and APIs, the manufacturer shares material-related data—such as carbon emissions—with the supplier in a standardized and interoperable format.
    c. To ensure trust and compliance, all interactions are validated by CSP-B using IAM. Additionally, the CAB certifies that the data exchange aligns with Catena-X standards, reinforcing trust between both parties.
    d. As a result of this collaboration, the manufacturer utilizes a Business Application provided by a Business Application Provider (BAP) to analyze the shared data. The insights gained from this analysis support both the manufacturer and the supplier in identifying concrete actions to reduce their carbon footprint.

5.2.4.2.3 Architecture Involvement

1. Development Environment: The KIT, developed by Eclipse Tractus-X, ensures the tools are open-source and standardized. Feedback from the battery manufacturer is sent back to *the Association* for refining the KIT or creating new standards.

2. Operating Environment

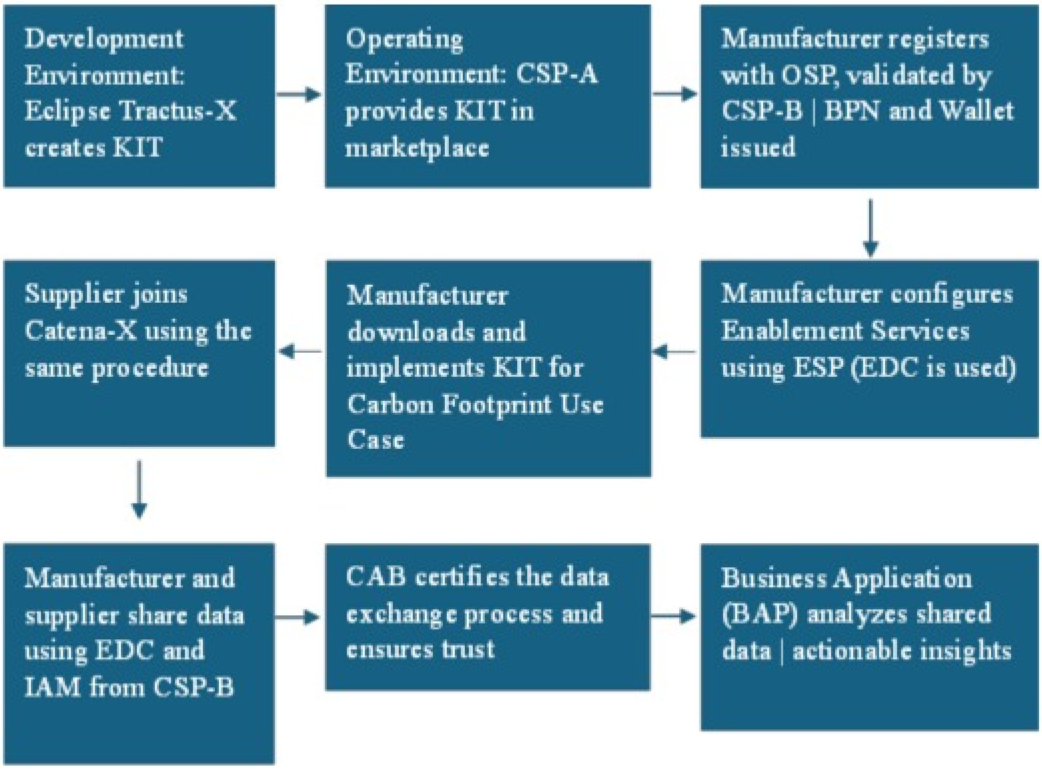
    - Enablement Services: Ensure connectivity between the manufacturer and supplier through the EDC.

    - Core Service A provides the marketplace for the KIT.

    - Core Service B manages IAM and trust services.

3. Governance and Certification: The Catena-X Association oversees the entire process, ensuring the rules are followed, while the CAB ensures that all participants remain compliant throughout their interactions.

5.2.4.2.4 Workflow



5.2.4.2.5 Outcome

By leveraging the Catena-X architecture, roles, and use cases, the manufacturer and supplier achieve transparency in data exchange, trust via certifications and IAM, actionable insights through standardized tools and applications, and sustainability with a reduced carbon footprint in their supply chain.

5.2.5 CATENA-X | DATA EXCHANGE BASED ON SSI

The self-sovereign identity (SSI) is a key component of the Catena-X ecosystem. It allows participants to control their own identity credentials without relying on centralized authorities for every interaction.

5.2.5.1 Role of CSP-B

CSP-B validates (or revokes) participants' requests for VCs and is currently the sole issuer of them in the Catena-X ecosystem. It ensures that all VCs meet the necessary standards and are interoperable within the ecosystem.

5.2.5.2 Gaia-X compliance

The SSI framework is integrated with the Gaia-X Clearing House for identity validation. It also ensures compliance with eIDAS (Electronic Identification and Trust Services) [28] regulations, enhancing the legal trustworthiness of issued credentials.

5.2.5.3 The data exchange process

As has been mentioned, the data exchange based on SSI in Catena-X begins with participants acquiring credentials from CSP-B, which validate their identity and compliance. Data providers then publish offers specifying access and usage policies, while data consumers query these offers and present their credentials for validation. Once validated, consumers review the provider's policies and, if acceptable, negotiate and finalize a contract. Data transfer occurs directly between endpoints via EDCs, with policies automatically enforced to ensure compliance. [25]

5.3 EONA-X

EONA-X is the mobility, transport and tourism data space [29]. Although created in 2022, it's kick-off was not until the 2024 Paris Olympic games [30], when it was implemented in order to have an adequate information transfer and control of the tourism in France.

EONA-X is the other data space that uses the EDC connector. Therefore, it has an architecture like the one Catena-X has, and the data transfer process is almost the same. To highlight the main difference in the adoption of data spaces between Catena-X and EONA-X, Catena-X can be considered an Innovator within the data spaces sector, and EONA-X an early adopter [31].

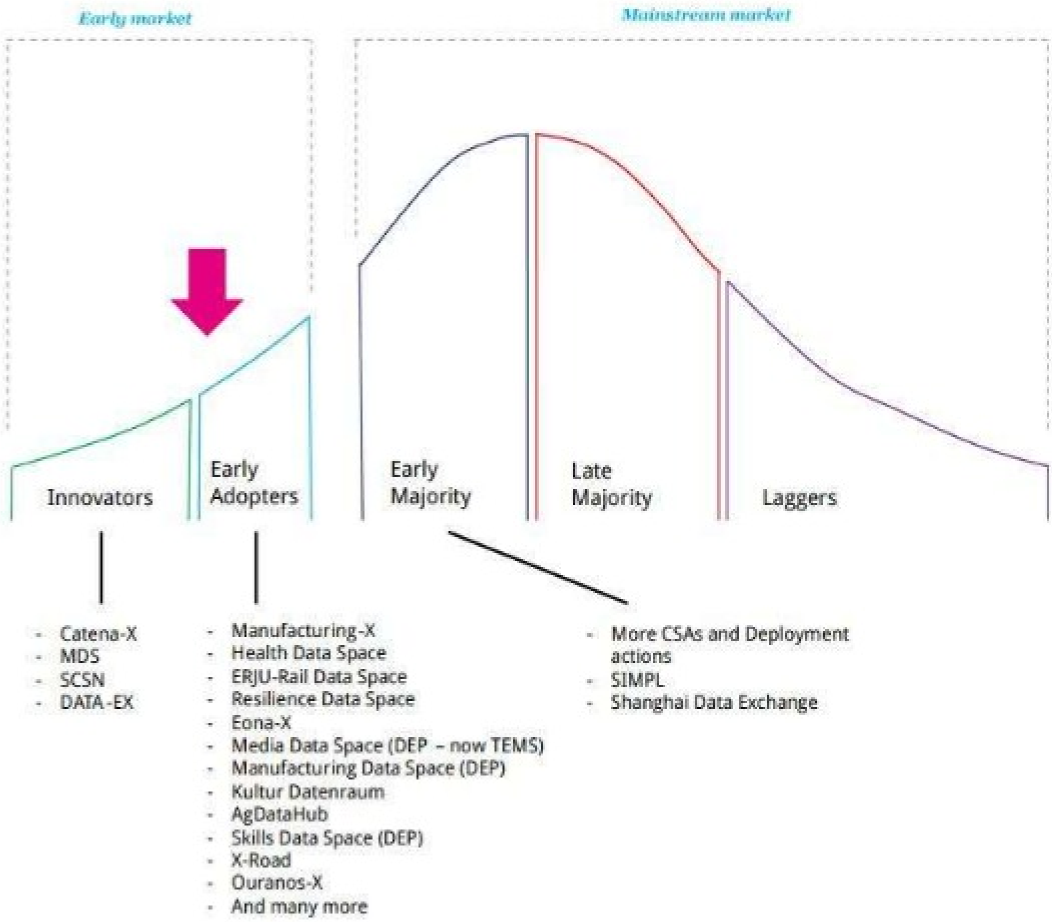


Figure 14. Adoption of data spaces with some examples [31]

Chapter 6. DEVELOPED MODEL

6.1 DATASPACES INTERCONNECTION PROPOSAL

Digital transformation has driven the adoption of sector-specific data spaces to facilitate collaboration among companies within a single ecosystem. As previously mentioned, projects such as Catena-X in the automotive industry, EONA-X in the mobility and tourism sector, and Omega-X in energy have demonstrated the viability of federated models for sharing information among companies in the same sector. However, these data spaces currently operate in isolation, which prevents a comprehensive view of the value chain and limits cross-sector innovation. Until now, if a company—such as Gestamp—wanted to transfer data with a company connected to EONA-X but not to Catena-X, Gestamp would need to connect to EONA-X, in addition to already being connected to Catena-X, in order to transfer data. This would be inefficient and would increase costs, resources, and time.

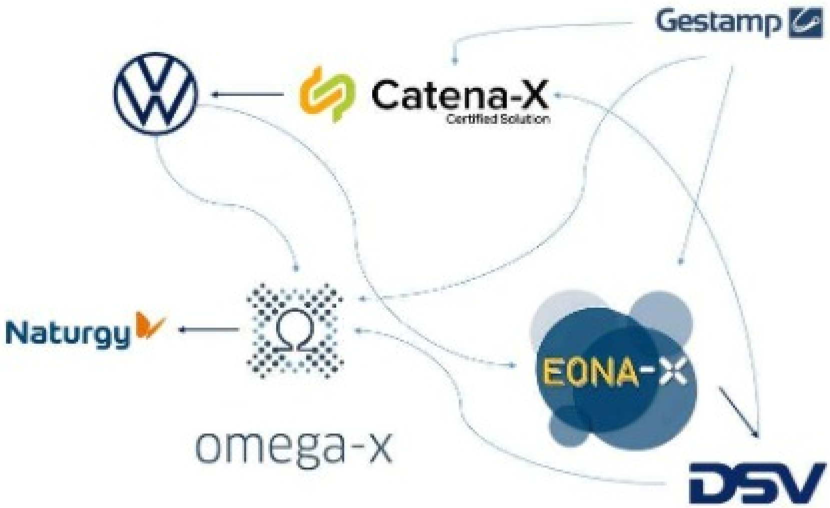


Figure 15. Currents status of communication among data sectors

Figure 15 shows how, currently, companies that want to consume data from entities in other data spaces will have to connect manually to the provider's data space, which will be independent to their own data space.

To address this issue, an interconnection infrastructure for data spaces is proposed under Gaia-X governance, with the goal of ensuring interoperability and security in information exchanges. In this model, Gaia-X serves as the central interconnection node between data spaces, providing the necessary standards and protocols for communication among them.

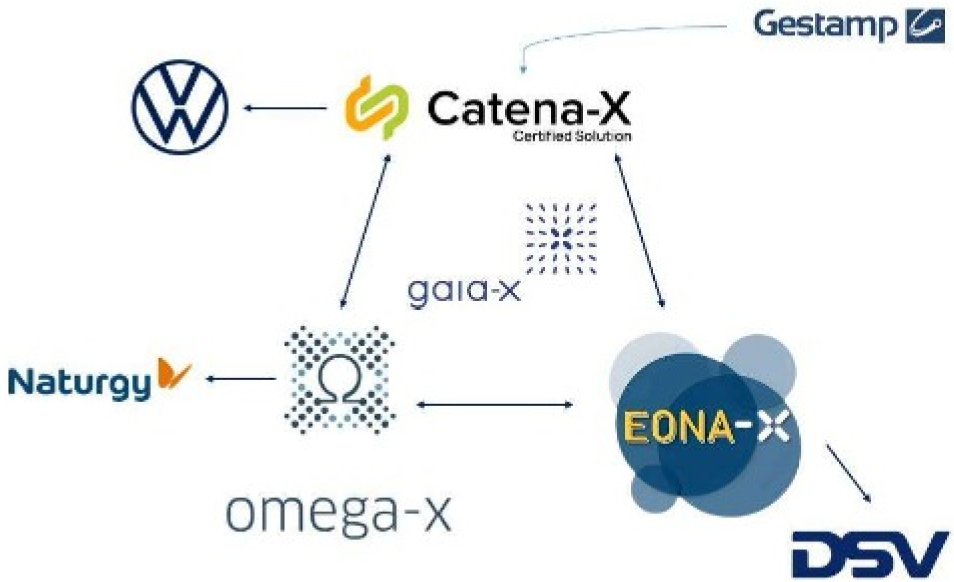


Figure 16. Interconnection proposal with Gaia-X as a central node

Figure 16 shows how, in the proposed interconnection model, Gaia-X will be serving as a central node which enables communication and interconnection within data spaces. This proposal is based on the creation of the Gaia-X Interconnectivity Hub, a connectivity layer that will enable communication among data spaces from different sectors without compromising their independence. This model follows an approach that is based on telecommunications network architectures.

The Interconnectivity Hub is based on a design with two main layers:

1. Governance and Security Layer:

Implemented by Gaia-X through its AISBL and the Gaia-X Digital Clearing House (GXDCH), it is responsible for certifying identities. Digital identity certification is carried out through Decentralized Identifiers (DIDs) and VCs. However, it is important to highlight that what have so far been considered identities were essentially companies. Now, each data space will have its own identity granted by Gaia-X, and each company within that data space will have a "child" identity within the data space (similar to how, in telecommunications networks, a network might have a generic IP, such as 10.0.0.0, but then an end device has the IP 10.0.0.1—something similar could happen with data spaces). These hierarchical identities could be managed by Gaia-X. Gaia-X would be the entity in charge of verifying, identifying, and validating each data space. This ensures that each data space meets the standards of interoperability and digital trust.

2. Federated Connectivity Layer (Interconnectivity Hub):

   It would be implemented with a network of semantic connectors based on the EDC. Each data space is connected to the Interconnectivity Hub through an Interconnectivity Hub Connector (IHC) - a tool that functions like the EDC but serves to connect data spaces to the Interconnectivity Hub. In other words, companies connect to the data space using an EDC, while the data spaces connect to the Interconnectivity Hub through the IHC. The IHC can be seen as an EDC for inter data space data transfers which enables connection to the Interconnectivity Hub.

The proposed interconnection model follows principles of IP architecture, where each data space behaves like a private network, and Gaia-X acts as an Internet Service Provider (ISP) that manages data routing between different ecosystems. Each data space assigns unique identifiers to its members, called BPNs. In this proposal, there are two types of BPNs:

- Private BPN: Enables communication within the same data space.
- Public BPN: Enables communication with external data spaces.

The Interconnectivity Hub would maintain a data space routing table, similar to how routers within an ISP manage IP addresses on the Internet. If an entity from Catena-X wants to communicate with an entity from EONA-X, the request is redirected through Gaia-X's Interconnectivity Hub, which acts as a trusted intermediary. In other words, each data space acts like a node (private network) in the larger network, and the Gaia-X Interconnectivity Hub functions as the intermediate ISP routers that route communication between different networks, and companies act as the end devices on the network connected to their respective data spaces (private networks). Companies connect to data spaces via the EDC, and data spaces connect to the Interconnectivity Hub using the IHC. This means that, within a data space, all companies have a unique EDC, but they all share a common IHC.

Nevertheless, it is important to recap on how telecommunication networks work.

6.1.1 TELECOMMUNICATIONS NETWORKS

First, it is important to consider the layered communication models: the OSI model and the TCP/IP model.

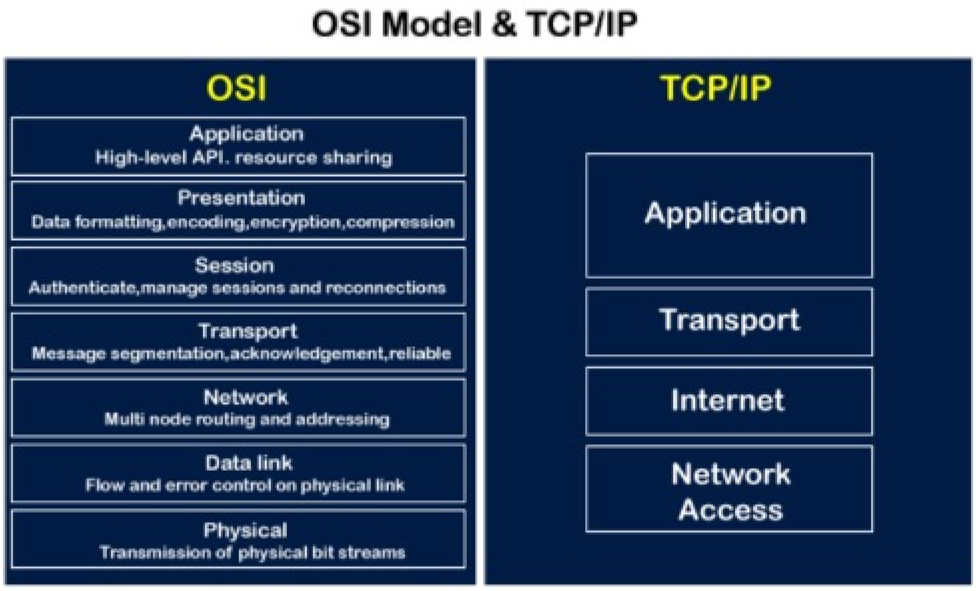


Figure 17. OSI Model and TCP/IP

6.1.1.1 OSI Model

In the OSI model, there are 7 layers:

1. Physical Layer: Responsible for the raw transmission and reception of bits through the physical medium (cables, fiber optics, wireless, etc.). It defines electrical and mechanical aspects, such as voltages, transmission rates, and types of connectors. It does not deal with frames or packets; it only transfers electrical or optical signals (bits) between devices.
2. Data Link Layer: Organizes bits into frames and controls access to the medium. It ensures basic error detection and correction and manages flow control between adjacent nodes. It includes protocols such as Ethernet or Wi-Fi and deals with MAC (Medium Access Control) addresses.
3. Network Layer: Handles routing of packets from the source to the destination across multiple networks. It defines logical addresses and chooses the appropriate route. Examples include IP (in the TCP/IP model) or IPX, as well as concepts like routing and address assignment.
4. Transport Layer: Provides a data flow between source and destination. It manages data segmentation and reassembly, error control, and end-to-end flow control. The most common protocols at this layer are TCP and UDP.
5. Session Layer: Manages and maintains sessions (dialogues) between two applications. It creates, controls, and ends the logical connections.
6. Presentation Layer: Transforms data into the format expected by the Application layer or the receiving system (e.g., encryption, compression, or format conversion). It deals with data encoding and security through encryption or the TLS/SSL protocol.
7. Application Layer: Provides services that user applications use to communicate. It includes high-level protocols such as HTTP, FTP, SMTP, etc. It is the layer closest to the user.

6.1.1.2 TCP/IP Model

In the TCP/IP model, there are 4 layers:

1. Network Access Layer: Handles the physical transmission of data and addressing on the local network. It includes link-layer encapsulation methods (Ethernet, Wi-Fi, etc.) and deals with MAC addresses.
2. Internet Layer: Manages logical addressing (IP) and the routing of packets.
3. Transport Layer: Provides end-to-end communication between processes, handling segmentation, flow control, and error recovery as needed. The key protocols are TCP and UDP.
4. Application Layer: Includes all application protocols (HTTP, FTP, SMTP, DNS, etc.). This is where services and direct interactions with the user or the final application take place.

Regarding the relationship between the OSI layers and the TCP/IP layers, in the OSI model, the Physical (1) and Data Link (2) layers are usually grouped together in the TCP/IP Network Access Layer. The Network (3) layer in the OSI model basically corresponds to the Internet layer in TCP/IP. Additionally, the Transport (4) layer in the OSI model is analogous to the Transport layer in TCP/IP. Finally, theSession (5), Presentation (6), and Application (7) layers in the OSI model are combined into the Application layer in the TCP/IP model.

6.1.1.2.1 End-to-end communications

A typical end-to-end communication scenario in a telecommunications network might look like this:

Suppose two users, Alice and Bob, want to communicate. The TCP/IP model will be the one chosen rather than the OSI model because it is more widely used in today's networks. Let's assume Alice has a private IP address of 192.168.1.10 and a public IP address of 85.120.34.56, while Bob has a private IP address of 192.168.2.20 and a public IP address of 102.45.67.89. Obviously, they are on different private networks.

The communication will undergo 9 steps:

1. Alice wants to communicate with Bob:

- For example, Alice wants to visit the web page hosted on Bob's server or send him a file.
- This starts at the Application layer (HTTP, FTP, etc.): Alice enters a URL in her browser or sends a file. By using the URL, Alice attempts to connect to Bob's public IP (e.g., 102.45.67.89). This causes the Application layer to "translate" and encode the user input to pass it to the Transport layer.

2. Transport Layer (TCP/UDP):
   - The first data segmentation takes place here. Additionally, source and destination ports are added (e.g., HTTP → 80, HTTPS → 443).
3. Network Layer and Routing:
   - The packet is encapsulated with IP addresses. The Network layer places the IP addresses: source address = Alice's private IP, destination address = Bob's public IP.
   - The packet is then sent to the Data Link layer.
4. Data Link Layer:
   - The frame is sent to the gateway (Router A).
   - Alice looks up the MAC address of her router (192.168.1.1) via ARP (Address Resolution Protocol).
   - The Data Link layer encapsulates the packet with Source MAC = Alice's MAC and Destination MAC = Router A's MAC.
   - Router A is called the border router because it sits between Alice's private network and the internet.
5. Router A processes the packet:
   - Router A performs address translation via NAT (Network Address Translation).
   - Because Alice's IP 192.168.1.10 is a private IP, the router translates it to a public IP (e.g., 85.120.34.56).
   - The source IP changes from 192.168.1.10 to 85.120.34.56.
   - Router A does not have a direct connection to Bob's network.

- Therefore, Router A sends the packet to its ISP (Internet Service Provider) so it can find Bob's network.

6. Routing on the Internet:
    - The ISP's routers consult their routing tables and forward the packet along.
    - To route correctly, these routers use protocols such as BGP (Border Gateway Protocol), which connects different Autonomous Systems (AS) and selects the best route on the internet; or OSPF (Open Shortest Path First), which finds the fastest route within an ISP or organization.
    - These routers are known as intermediate routers.
    - It's important to note that throughout the routing process, the destination IP remains Bob's public IP, not his private IP.
7. Arrival at the destination:
    - The intermediate routers deliver the packet to the border router on Bob's private network (Router B).
    - Router B performs Reverse NAT (or inbound NAT). It translates Bob's public IP 102.45.67.89 to his private IP 192.168.2.20.
8. Delivery to Bob:
    - Data Link Layer:
      The border router looks up Bob's MAC address using ARP. Then, the packet is sent to Bob's MAC address.
    - Network Layer:
      The packet is delivered to Bob's private IP address.
    - Transport Layer
      Bob's ports are established to carry out the communication. Additionally, and only ff necessary, TCP confirms that the message has been received.
    - Application Layer
      Bob receives the message from Alice.
9. Bob's response
    - The process is reversed.

- The source IP will now be 192.168.2.20.
- The destination IP will be 85.120.34.56.
- The routers carry out NAT and routing until the response reaches Alice.

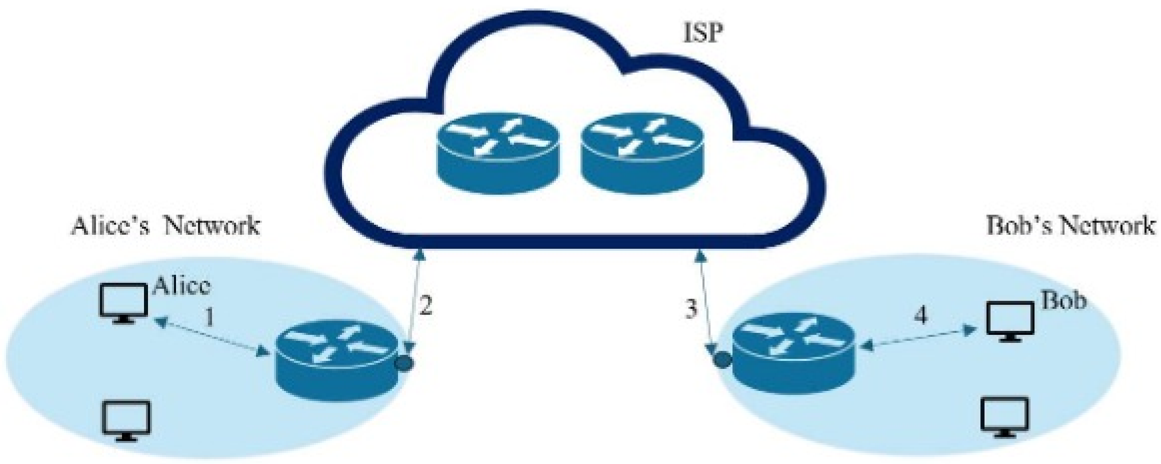


Figure 18. Simplified telecommunications end-to-end communication

In Figure 18, a simplified telecommunications end-to-end communication is shown. There are four main steps within the communication. Each of the steps will contain a frame, with a source IP and a destination IP address.

User	Public IP	Private IP
Alice	85.120.34.56	192.168.1.10
Bob	102.45.67.89	192.168.2.20

Table 3. Alice's and Bob's public and private IPs

Communication	Source IP	Dest. IP	Packet description
1	192.168.1.10	102.45.67.89	A packet is sent from Alice's private IP to Bob's public IP.

2	85.120.34.56	102.45.67.89	NAT is made in the border router A. A packet is sent from Alice's public IP to Bob's public IP.
3	85.120.34.56	102.45.67.89	Although the packet traverses multiple ISP routers on its way to Bob's network, the source and destination IP addresses in the Network layer header remain unchanged. However, the MAC addresses at the Data Link layer do change from one hop to the next.
4	85.120.34.56	192.168.2.20	NAT is performed by the border router B. A packet is sent from Alice's public IP to Bob's private IP. This is known as port forwarding.

Table 4. End-to-end source and destination IP changes

In Table 4, the changes in the origin and destination IPs throughout the hops is described.

6.1.2 Extrapolation of Telecommunications Network Architectures to Inter Data Spaces Communication Proposal

After examining on a large scale how end-to-end communications work in telecommunication networks, this section uses that as an analogy to propose how communication among data spaces (inter-data spaces data transfers) could work.

6.1.2.1 Gaia-X Accreditation of Data Spaces and Assignment of Public BPN

Using DIDs (Decentralized Identifiers) and VCs (Verifiable Credentials), Gaia-X identifies and accredits each data space, assigning it a public BPN.

Suppose EONA-X applies to be a Gaia-X-accredited data space. Gaia-X verifies EONA-X as an official data space and grants it a public BPN. For this example, let's say it is "101011". Once EONA-X has its public BPN, it can connect to the Interconnectivity Hub through the IHC that enables communication between different data spaces (explained later).

6.1.2.2 Connecting Companies to a Data Space and Assigning a Private BPN

Let's suppose the fictitious company "Transport S.L." wants to connect to EONA-X. Since EONA-X is already accredited by Gaia-X, it can issue a credential (VC) and a private BPN within its data space to "Transport S.L." Let's assume this private BPN is "0001".

Thus, "Transport S.L." becomes uniquely identified within EONA-X by concatenating the space's public BPN ("101011") with its own private BPN ("0001"), resulting in 101011:0001 as the complete identifier.

Once "Transport S.L." has its private BPN, it can connect and exchange data within EONA-X via its EDC. In addition, by being accredited within EONA-X, it can use EONA-X's IHC to communicate with other data spaces. Therefore, each data space's public BPN is analogous to a public IP address in telecom networks, while a company's private BPN is analogous to a private IP. When communicating across data spaces, traffic goes through the Interconnectivity Hub via IHC, which acts like a trusted router or intermediary under Gaia-X governance.

6.1.2.3 Catena-X Accreditation and Assignment of Its Public BPN

Catena-X follows the same process: Gaia-X accredits Catena-X as a valid data space and assigns it a public BPN, for example, "202021". Catena-X thus obtains its own IHC, which all members of Catena-X can use to communicate with other data spaces.

Then, a company like Gestamp can connect to Catena-X and receive a private BPN, let's say "0005". Its complete identifier would be 202021:0005 (public BPN of Catena-X : private BPN of Gestamp).

Gestamp connects to its data space (Catena-X) using an EDC and can leverage Catena-X's IHC to communicate with other nodes such as EONA-X.

6.1.2.4 End-to-End Communication Example Between Companies in Different Spaces

Once all companies (e.g., "Transport S.L." and Gestamp) have their own private BPNs in their respective data spaces (EONA-X and Catena-X), we can illustrate how an end-to-end communication occurs when Gestamp wants to contact a company outside its space (in this case, "Transport S.L.").

Gestamp (full BPN 202021:0005), connected to Catena-X via an EDC, attempts to communicate with "Transport S.L." (101011:0001). If Gestamp wanted to contact another company in Catena-X, it would only need the private BPN (e.g., 0007). But because 101011:0001 is not found within Catena-X, this is assumed to be an inter-data-space communication.

6.1.2.4.1 Gestamp activates the IHC

Using Catena-X's IHC, Gestamp contacts the Interconnectivity Hub via the IHC to send a Contract Request to the BPN 101011:0001.

The Gaia-X Interconnectivity Hub identifies the destination, as it belongs to Gaia-X, and Gaia-X maintains a registry of the public BPNs for all data spaces (since Gaia-X assigned them in the first place). Recognizing 101011, it knows that it belongs to EONA-X and forwards the Contract Request to EONA-X's IHC. Since "Transport S.L." has the full BPN 101011:0001, EONA-X locates that company and delivers the Contract Request. A pinhole is opened between Gestamp and "Transport S.L.". They negotiate the conditions of their data exchange through their IHCs (in the same way as a contracts are negotiated through EDC in intra data spaces communications) and once agreed, a data exchange can begin. With this channel active, both parties exchange data in the same way as if they were in a single data space, but using the Interconnectivity Hub as a "bridge" between Catena-X and EONA-X.

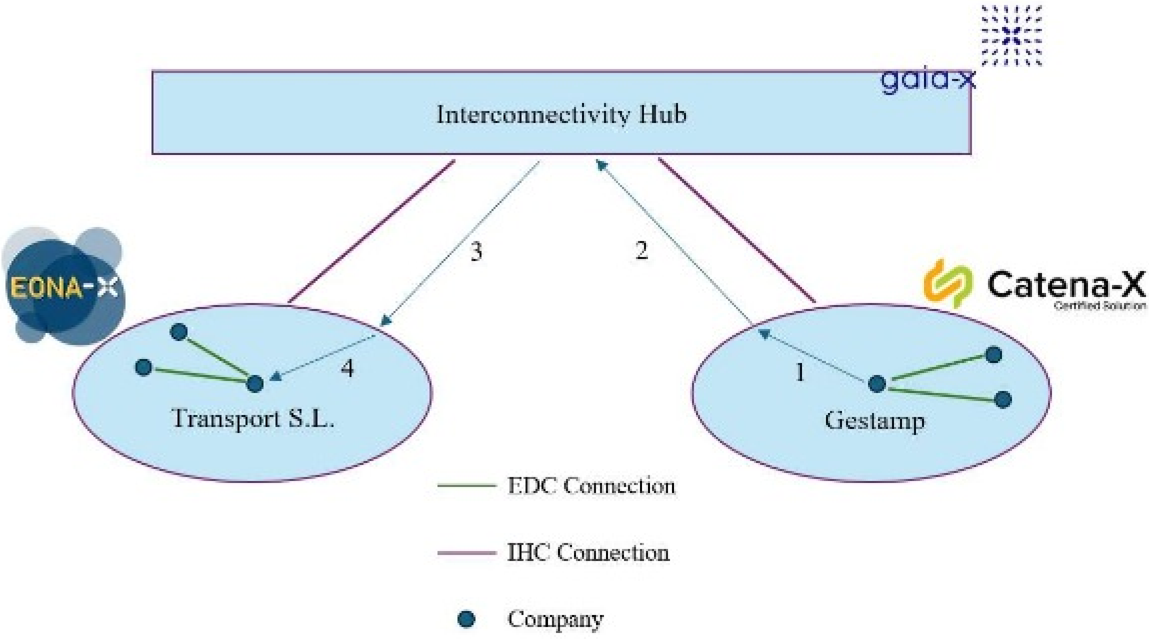


Figure 19. End-to-end inter data space communication proposal

Figure 19 shows the proposed inter data space communication.

Step	Description
1	Gestamp (BPN 202021:0005) makes a contract request with entity with BPN 101011:0001 via EDC. Since the public BPN differs from that of the current data space, inter data-space communication must take place.. IHC is activated.
2	Contract Request is forwarded to Interconnectivity Hub. The Interconnectivy Hub, thanks to its belonging to Gaia-X, can access the registry of all public BPNs mapped to their corresponding data space.
3	Public BPN 101011 corresponds to EONA-X. Contract Request is forwarded to this data space.
4	Contract Request is forwarded to Transport S.L. BPN.

	Contract Request is received by Transport S.L. If the negotiation is successful, Gestamp and Transport S.L. can transfer data via the IHC, always through the opened pinhole of the Interconnectivity Hub which acts as a proxy

Table 5. Inter data space communication steps

Table 5 summarises the different steps that take place in the communication proposal.

6.2 DATASPACES INTERCONNECTION ANALYSIS

The proposed architecture introduces a number of characteristics that make it technically suitable to allow data transfer among data spaces. From an engineering point of view, it builds upon well-established protocols and paradigms in telecommunications, minimizing the need for heavy innovation. In other words, the proposed model would undergo a brownfield project rather than a greenfield one, redeveloping an existing infrastructure saving on initial costs.

On the one hand, the implementation complexity is not exhaustive since the proposal does not rely on inventing entirely new disruptive technologies. Rather, it adapts well-established concepts from the telecommunications networks such as public/private IP addressing, NAT routing, autonomous systems, Internet Service Providers (ISPs), and routing mechanisms. This reuse of protocols reduces the R&D overhead compared to the alternative of building a network from scratch. In the same line, the IHC would be built upon an extension of the EDC, preserving its control and data plane (seen later), but reconfigured to support inter-data space interactions rather than solely intra-data space communications.

Furthermore, the Interconnectivity Hub introduces a valuable governance and interoperability layer. It credentials all participants through VCs, aligning with the Gaia-X vision where all entities have to be certified.

However, from a systems architecture perspective, the introduction of a central component brings the risk of centralization and single points of failure. Positioning the Interconnectivity Hub as a neutral interconnection point creates a choke point in the architecture. Its failure could lead to an entire inter-space communication failure. To mitigate this, fault tolerance and high availability must be explicitly addressed—e.g., through redundancy, load balancing, and distributed control mechanisms.

Additionally, Gaia-X is intrinsically federated in its design and is not intended to operate with centralized infrastructure components. Therefore, the development of the Interconnectivity Hub must not lead to a technical or political monopoly. A viable de-centralized approach would involve federating the Interconnectivity Hub into multiple certified nodes, much like the architecture of DNS root servers [32]. These nodes would be distributed and independently governed.

Another architectural consideration involves the use of a concatenated BPN public-private identifier schema as a globally unique identifier (GUID) within data spaces. For this system to scale, it will require an infrastructure based on Distributed Hash Tables (DHTs) [33] or DNSSEC-like [34] systems. Such mechanisms also prevent identifier collisions, thereby enforcing global uniqueness.

A number of Quality of Service (QoS) parameters must be guaranteed across the network infrastructure. These include:

- Low latency: particularly important for real-time or near-real-time data services.
- Minimum guaranteed throughput: to avoid congestion.
- Resilience and redundancy mechanisms: including failover paths and route diversity.

These technical requirements ensure that data exchange between interconnected data spaces remains robust, sovereign, secure, and trustworthy, even under fault.

6.3 DATASPACE CONNECTIVITY | EDC

In the previous sections, the Eclipse Dataspace Connector (EDC) has been discussed as the foundational technology for the development of the IHC. However, it is crucial to gain a deeper understanding of how the EDC operates, what components it comprises, how it is used, and how it can be practically implemented. In other words, it is appropriate at this stage to delve into the EDC technology in more detail and even attempt a data exchange within Catena-X as a first step toward the future development of the IHC.

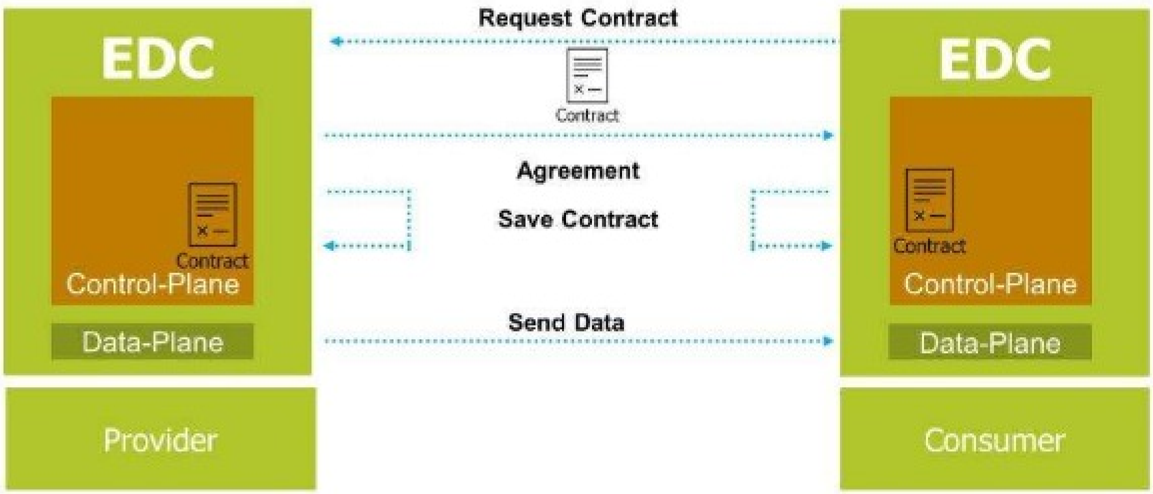


Figure 20. EDC componets [35]

The EDC, Figure 20, is an open-source technology developed to enable data exchange between data spaces through formal contracts and under sovereign control [36]. For data exchange to occur using this technology, both the data provider and the data consumer must be integrated into the data space using an EDC instance.

The EDC facilitates the direct transmission of data from one endpoint to another using standardized formats. It does not rely on intermediary databases, external services, or third-party providers. Each EDC instance—on both the provider and consumer sides—contains two main components: the Control Plane and the Data Plane.

The Control Plane manages the orchestration of the EDC. It handles contract negotiation, data access policies, coordination of transfers, and authentication and authorization of the parties involved.

The Data Plane executes the actual data transfer. It supports multiple data exchange protocols and enables high-throughput transmissions.

The data exchange process between two EDC-integrated parties follows these steps:

Provider Side:

1. The provider creates a Data Asset.
2. The provider defines a Policy, indicating which consumer is allowed to access the associated data.
3. The provider creates a Contract Definition linking the Data Asset to the Policy. This defines that the consumer specified in the policy is permitted to access the Data Asset.
4. The provider publishes the Contract Definition.

Consumer Side:

5. The consumer queries the provider's catalog.
6. The consumer sends a Contract Request (initiates negotiation with provider by requesting it's data)

Provider Side:

7. The provider accepts the Contract Request.
8. The provider transfers the data to the consumer.

Steps 1 through 7 take place entirely within the Control Plane. Only step 8—the transmission and receipt of data—is handled by the Data Plane. The Control Plane governs the preconditions for the exchange; the Data Plane performs the exchange.

As previously mentioned, it is the Eclipse Tractus-X project the one that provides the reference implementation for the EDC within the Catena-X. While this implementation has been the most studied and developed in the context of this project, other deployments of the EDC exist that are not tied to Catena-X[2] and are intended for more general-purpose scenarios.

6.3.1 BASE EDC PROJECT

The principal Eclipse Data Connector (EDC) project is located in the repository https://github.com/eclipse-edc/Connector. It provides an extensible framework for building data space connectors that comply with the Dataspace Protocol (DSP) by IDSA and Gaia-X specifications. It serves as a generic and adaptable foundation for multiple use cases across various industries. Consequently, the Tractus-X EDC can be considered an adaptation of this base connector within Catena-X.

This repository contains the core modules required to build a functional connector, including:

- TransferProcessManager, ProvisionManager, and DataFlowManager, which handle data transfer processes and provisioning.
- SPI (Service Provider Interfaces), which allow for extensibility and customization of the connector's behavior.
- Launchers, which are executable packages that combine these modules into a deployable application.
- Protocol implementations for data exchange, such as the DSP.

[2] The **Eclipse** Dataspace Connector (EDC) is hosted by the **Eclipse** Foundation. Tractus-X uses a tailored version of the EDC adapted to the requirements of the Catena-X ecosystem. Other domains may also use the EDC, slightly different to the Tractus-X EDC.

The Catena-X connector inherits these core components from the base EDC and integrates additional features required for operation within the automotive ecosystem. That is why it is known to be an extension of the EDC.

To deploy a basic version of the EDC and prepare it for interaction with other connectors, the following steps can be followed:

1. Clone the repository.

2. Build the EDC using the command: ./gradlew.bat launchers:sts-server:build

3. Deploy the connector using: java -Dedc.fs.config=launchers/sts-server/config.properties -jar launchers/sts-server/build/libs/sts-server.jar

```
web.http.port=8181
web.http.path=/api
web.http.sts.port=9292
web.http.sts.path=/api/v1/sts
edc.iam.sts.clients.first.name=Test Client
edc.iam.sts.clients.first.id=testClientId
edc.iam.sts.clients.first.client_id=testClient
edc.iam.sts.clients.first.did=did:example:first
edc.iam.sts.clients.first.secret.alias=secretAlias
edc.iam.sts.clients.first.private-key.alias=private-key
edc.iam.sts.clients.first.public-key.reference=public-key
edc.sts.server.vaults.private.key=private-key
edc.sts.server.vaults.private.value=-----BEGIN PRIVATE KEY-----\r\nMIIEvQIBADANBgkqhkiG9w0BAQEFAASCBKcwggSjAg
edc.sts.server.vaults.secret.key=secretAlias
edc.sts.server.vaults.secret.value=clientSecret
```

Figure 21. Generic EDC deployment

4. Generate an authentication token using:

   curl --request POST --url http://localhost:9292/api/v1/sts/token \

   --header "Content-Type: application/x-www-form-urlencoded" \

   --data grant_type=client_credentials \

   --data client_id=testClient \

   --data client_secret=clientSecret \

   --data audience=test10

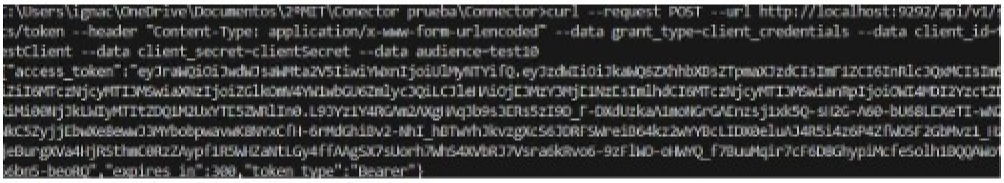


Figure 22. Generic EDC token generation

The token generated in Step 4 is used to authenticate and authorize communication between this EDC instance and other EDC endpoints. This token is essential, as data exchange via EDC requires secure authentication and is not open by default. At this stage, the connector is built, deployed, and ready to participate in authenticated data exchanges with other EDCs.

However, this setup does not yet support actual data exchange. To enable it, the type of data and the sharing logic must be explicitly defined—something that is pre-configured in the Tractus-X EDC to align with Catena-X use cases. Additionally, data-sharing policies and conformance to relevant data space standards must be established: this default connector only represents a generic, abstract implementation of the EDC. It is intended for foundational understanding and for adaptation to specific data spaces but is not suitable for demonstrating a complete data exchange workflow.

For that purpose, a more specific implementation is needed. Therefore, the following sections will analyze data exchange using the Tractus-X EDC, which includes preconfigured components aligned with Catena-X requirements. Within the Tractus-X community, additional repositories have also been developed to simplify deployment and configuration of EDC connectors.

6.3.2 TRACTUS-X PRINCIPAL PROJECT

The Tractus-X EDC main project can be found in this repository: https://github.com/eclipse-tractusx/tractusx-edc. It is the foundational component of the Catena-X EDC infrastructure, and inherits many components mentioned in the 6.3.1 section.

This repository is maintained by the Eclipse Tractus-X working group and includes predefined APIs to make policy management mechanisms, data transfers and extensions for handling digital twins. However, at the moment of deploying this connector, some issues were faced.

First, the repository does not offer a ready-to-use dataspace environment. Instead, it serves as a low-level building block to create and configure individual EDC instances. For practical testing and validation, such as simulating data exchange between participants, it is necessary to deploy at least two EDCs—one acting as a data provider, and the other as a consumer. This repository requires significant manual configuration and orchestration of auxiliary services making it unsuitable as a testing tool. Therefore, it is typically not used directly on its own, but rather as a dependency in higher-level orchestration tools like the MXD Chart and the Umbrella Chart, which are explored in the following sections. These deployment charts offer intuitive onboarding-oriented environments, enabling faster local testing.

In addition, a critical compatibility issue was identified during the evaluation phase. The repository currently builds and deploys version 0.10.0 of the EDC. However, the Cofinity-X ecosystem only supports up to version 0.7.0. As such, even if an EDC were successfully deployed using this repository, it would not be eligible for integration into the Cofinity-X ecosystem due to version mismatch. A previous version deployment would be needed. This misalignment between the repository version and the Cofinity-X environment, among the other reasons explained, prevented further use of this EDC instance.

Instead, work on other charts has been performed.

6.3.3 TRACTUS-X MXD CHART

In order to deploy a Minimum Tractus-X Dataspace "MXD" locally, the Tractus-X community developed the following repository: https://github.com/eclipse-tractusx/tutorial-resources/blob/main/mxd

The purpose of the MXD repository is to simulate a data transfer between two participants, Alice and Bob, by creating two EDC instances. Additionally, the setup involves the exchange of data stored in PostgreSQL databases between the two simulated users.

The deployment of the MXD chart involves the following steps:

1. Cluster creation: A Kubernetes cluster must be provisioned using Docker Desktop. Running a dataspace with two EDC instances requires orchestration of several containers, each composed of multiple distributed components—specifically, two EDC connectors (Alice and Bob), a central catalog service, and two Vault[3] instances (one per participant), two PostgreSQL databases (one per participant), and one Identity Hub [4] instance.

Figure 23. Verification that the MXD cluster is created

2. Create runtime image: After the cluster and its components are provisioned, a Docker runtime image is built. This image encapsulates the complete environment required to run the various components of the system.
3. Deploy all cluster resources: Use Terraform to deploy all necessary resources within the cluster. Terraform automates the provisioning and configuration of the infrastructure components.
4. Verify deployment: Once the resources have been deployed, the Terraform output can be reviewed to verify that the relevant components are running. This includes the data plane and control plane for both Alice and Bob, as well as the PostgreSQL databases for each user.

---

[3] Vault: Secret tool developed by HashiCorp [35].
[4] Identity Hub: A service that manages decentralized data and enables the secure exchange of Verifiable Credentials [36].

```
alice-database-credentials = {
  "database-host" = "10.96.121.67"
  "database-port" = 5432
  "database-url" = "10.96.121.67:5432"
  "instance-name" = "bob"
}
alice-node-ip = "10.96.142.198"
alice-urls = {
  "health" = "http://localhost/alice/health"
  "management" = "http://localhost/alice/management/v3"
  "proxy" = "http://localhost/alice/proxy"
  "public" = "http://localhost/alice/api/public"
}
bob-database-credentials = {
  "database-host" = "10.96.84.53"
  "database-port" = 5432
  "database-url" = "10.96.84.53:5432"
  "instance-name" = "alice"
}
bob-node-ip = "10.96.192.244"
bob-urls = {
  "health" = "http://localhost/bob/health"
  "management" = "http://localhost/bob/management/v3"
  "proxy" = "http://localhost/bob/proxy"
  "public" = "http://localhost/bob/api/public"
}
```

Figure 24. Alice's and Bob's resources deployment

The next step involves testing data exchange and contract negotiation between the two users, Alice and Bob, using Postman. However, this phase presented unexpected difficulties. Although a contract request was initiated from one of the EDCs, the other EDC did not receive it. As a result, the provider was unable to acknowledge the Contract Request, preventing the establishment of a contractual agreement and, consequently, any data exchange.

It is important to note that the repository used for this setup had not been maintained regularly. At the time of implementation, it had not been updated for over 12 months and did not reflect the latest policies and protocols of the Tractus-X EDC protocols. Due to this lack of updates, the repository was no longer aligned with the current state of the platform and was deprecated by the Tractus-X community. This obsolescence was the root cause of the failure to establish a functional data exchange in the MXD setup. The repository contained outdated and misconfigured components that were incompatible with the current EDC specifications.

While the work on MXD provided an opportunity to understand the basic architecture and operation of the Tractus-X EDC, it is worth highlighting that the obsolescence of the repository was not communicated in advance. From Gestamp's perspective, this lack of prior notice led to a misallocation of effort and resources toward a setup that ultimately could not deliver a viable minimum dataspace.

The Tractus-X community now recommends using the Umbrella Chart repository, which will be discussed in the following section.

6.3.4 TRACTUS-X UMBRELLA CHART

The Umbrella Chart is a Helm chart [5] designed to orchestrate the deployment of all the components of the Tractus-X EDC, for users who want to test a production-like Tractus-X dataspace locally (e.g., on a Kubernetes cluster using Minikube) or in cloud environments. In this project, Minikube has been be used. The repository can be found in https://github.com/eclipse-tractusx/tractus-x-umbrella

Unlike older resources such as the MXD chart, the Umbrella Chart is actively maintained by the Tractus-X community.

The deployment of the Umbrella chart involves the following steps:

1. Cluster creation: A Kubernetes cluster is created with 4 CPU cores and 6 GB of RAM using Docker Desktop.
2. Download the chart dependencies of the umbrella helm chart.

[5] Helm chart: A package that bundles all the YAML files needed to deploy an application on a Kubernetes cluster. [37]

3. Download the 'Data Exchange' helm subset which enables data sharing between partcipants. This time, we will deploy three EDCs to simulate a communication between a provider and two consumers.
4. Once all pods are ready, data exchange with the API can begin using Postman.

The whole deployment of the Umbrella Chart, creating all the components needed to use both EDCs (both for the data provider and the data consumer) can be seen in Figure 25, Figure 26, and Figure 27.

Figure 25. Dashboard to visualize status of Deployments, Jobs, Pods and Sets of the two EDCs

```
NAME                                                        READY   STATUS             RESTARTS        AGE
pod/edc-dataconsumer-1-vault-0                              1/1     Running            0               29m
pod/edc-dataconsumer-2-vault-0                              1/1     Running            0               29m
pod/edc-dataprovider-vault-0                                1/1     Running            0               29m
pod/ssi-dim-wallet-stub-5597c94957-7rt5h                    1/1     Running            4 (21m ago)     29m
pod/umbrella-dataconsumer-1-db-0                            1/1     Running            1 (110s ago)    29m
pod/umbrella-dataconsumer-1-edc-controlplane-c96d9bc85-k6p4j 0/1    Pending            0               29m
pod/umbrella-dataconsumer-1-edc-dataplane-7754fdfcbf-bzlss  0/1     CrashLoopBackOff   7 (99s ago)     29m
pod/umbrella-dataconsumer-1-post-install-vault-setup-nv585  0/1     Completed          0               29m
pod/umbrella-dataconsumer-2-db-0                            1/1     Running            1 (82s ago)     29m
pod/umbrella-dataconsumer-2-edc-controlplane-664fb77c56-sszxf 1/1   Running            3 (24m ago)     29m
pod/umbrella-dataconsumer-2-edc-dataplane-75fcf6554f-qbfbp  1/1     Running            3 (22m ago)     29m
pod/umbrella-dataconsumer-2-post-install-vault-setup-c5tj7  0/1     Completed          0               24m
pod/umbrella-dataprovider-db-0                              1/1     Running            1 (110s ago)    29m
pod/umbrella-dataprovider-dtr-6d4889cf5c-zfmnt              1/1     Running            3 (111s ago)    29m
pod/umbrella-dataprovider-dtr-db-0                          1/1     Running            1 (110s ago)    29m
pod/umbrella-dataprovider-edc-controlplane-dcd699998-lhgm9  0/1     Pending            0               29m
pod/umbrella-dataprovider-edc-dataplane-85f68ddcd4-v8rsd    0/1     CrashLoopBackOff   7 (19s ago)     29m
pod/umbrella-dataprovider-post-install-testdata-2g67h       0/1     Error              0               21m
pod/umbrella-dataprovider-post-install-testdata-2pvxl       0/1     Error              0               18m
pod/umbrella-dataprovider-post-install-testdata-b6zdp       0/1     Error              0               2m25s
pod/umbrella-dataprovider-post-install-testdata-hbsqc       0/1     Error              0               23m
pod/umbrella-dataprovider-post-install-testdata-pfr8j       0/1     Error              0               21m
pod/umbrella-dataprovider-post-install-testdata-v57wd       0/1     Error              0               20m
pod/umbrella-dataprovider-post-install-vault-setup-cg8cj    0/1     Completed          0               24m
pod/umbrella-dataprovider-submodelserver-84bbf578b5-s5vfz   1/1     Running            2 (113s ago)    29m
pod/umbrella-pgadmin4-59bd698dd6-8m64x                      1/1     Running            0               29m
```

Figure 26. Console output of the EDC's Pods status

```
NAME                                                    READY   UP-TO-DATE   AVAILABLE   AGE
deployment.apps/ssi-dim-wallet-stub                     1/1     1            1           29m
deployment.apps/umbrella-dataconsumer-1-edc-controlplane  0/1   1            0           29m
deployment.apps/umbrella-dataconsumer-1-edc-dataplane   0/1     1            0           29m
deployment.apps/umbrella-dataconsumer-2-edc-controlplane  1/1   1            1           29m
deployment.apps/umbrella-dataconsumer-2-edc-dataplane   1/1     1            1           29m
deployment.apps/umbrella-dataprovider-dtr               1/1     1            1           29m
deployment.apps/umbrella-dataprovider-edc-controlplane  0/1     1            0           29m
deployment.apps/umbrella-dataprovider-edc-dataplane     0/1     1            0           29m
deployment.apps/umbrella-dataprovider-submodelserver    1/1     1            1           29m
deployment.apps/umbrella-pgadmin4                       1/1     1            1           29m

NAME                                                              DESIRED   CURRENT   READY   AGE
replicaset.apps/ssi-dim-wallet-stub-5597c94957                    1         1         1       29m
replicaset.apps/umbrella-dataconsumer-1-edc-controlplane-c96d9bc85   1      1         0       29m
replicaset.apps/umbrella-dataconsumer-1-edc-dataplane-7754fdfcbf  1         1         0       29m
replicaset.apps/umbrella-dataconsumer-2-edc-controlplane-664fb77c56  1      1         1       29m
replicaset.apps/umbrella-dataconsumer-2-edc-dataplane-75fcf6554f  1         1         1       29m
replicaset.apps/umbrella-dataprovider-dtr-6d4889cf5c              1         1         1       29m
replicaset.apps/umbrella-dataprovider-edc-controlplane-dcd699998  1         1         0       29m
replicaset.apps/umbrella-dataprovider-edc-dataplane-85f68ddcd4    1         1         0       29m
replicaset.apps/umbrella-dataprovider-submodelserver-84bbf578b5   1         1         1       29m
replicaset.apps/umbrella-pgadmin4-59bd698dd6                      1         1         1       29m

NAME                                            READY   AGE
statefulset.apps/edc-dataconsumer-1-vault       1/1     29m
statefulset.apps/edc-dataconsumer-2-vault       1/1     29m
statefulset.apps/edc-dataprovider-vault         1/1     29m
statefulset.apps/umbrella-dataconsumer-1-db     1/1     29m
statefulset.apps/umbrella-dataconsumer-2-db     1/1     29m
statefulset.apps/umbrella-dataprovider-db       1/1     29m
statefulset.apps/umbrella-dataprovider-dtr-db   1/1     29m

NAME                                                  STATUS     COMPLETIONS   DURATION   AGE
job.batch/umbrella-dataconsumer-1-post-install-vault-setup   Complete   1/1    4m55s      29m
job.batch/umbrella-dataconsumer-2-post-install-vault-setup   Complete   1/1    14s        24m
job.batch/umbrella-dataprovider-post-install-testdata        Running    0/1    23m        23m
job.batch/umbrella-dataprovider-post-install-vault-setup     Complete   1/1    16s        24m
```

Figure 27. Console output of the EDC's Deployments, Sets and Jobs status

However, a number of components fail to deploy correctly.

The following pods enter an error state:

- Data Plane of consumer 1
- Control Plane of consumer 1
- Data Plane of the provider
- Components related to the test data that the provider is expected to expose to the consumer

The failing deployments and replica sets correspond to:

- Data Plane of consumer 1
- Control Plane of consumer 1
- Data Plane of the provider
- Control Plane of the provider

In addition, the Job responsible for initializing the provider's test data also fails.

These issues indicate that both the data plane and control plane of the provider and consumer 1 fail to start correctly. As a result, no data exchange or communication can be established between the participants.

Following consultation with the external company *I.*—specialized in data connectors and data spaces—it was concluded that the issues encountered are due to hardware limitations, in contrast with the resource requirements specified in the official documentation of the Tractus-X Umbrella Chart repository.

According to the documentation, the Kubernetes cluster can be initialized with the following command:

minikube start --cpus=4 --memory=6gb

Even though it is not explicitly stated as so in the repository, this command corresponds to the minimum configuration.

Recommended configuration for a stable deployment:

- CPU: 6–8 vCPUs
- RAM: 10–12 GB

However, the local host machine does not meet these hardware requirements. This limitation has been identified as a blocking factor, as the available system resources are insufficient to successfully deploy the necessary components. As a result, critical pods failed to initialize, and the desired deployments could not be completed.

Given that local deployment is not feasible—neither with the generic EDC, nor with the base Tractus-X EDC, nor with the MXD Chart repository, nor with the Umbrella Chart—the decision was made to contract an *EDC as a Service* offering. This service was procured via the Cofinity-X Marketplace, selecting a provider that offers a fully managed EDC deployment.

6.3.5 TRACTUS-X EDC AS A SERVICE

As previously mentioned, an external EDC service was contracted. This approach delegates the entire deployment and orchestration of the connector to a third-party provider.

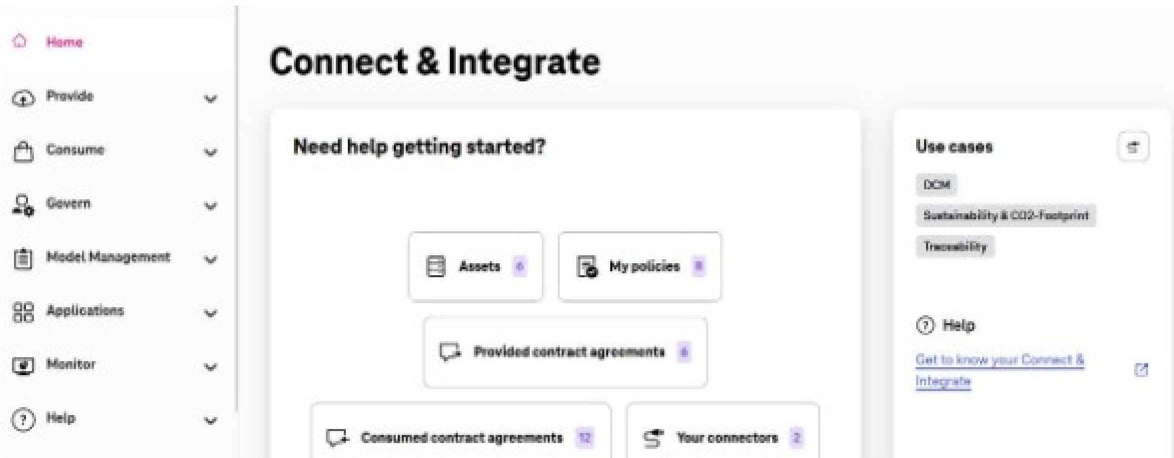


Figure 28. External platform's UI to orchestrate the EDC

With the connector now externally managed, the integration into Catena-X is successful. The only task remaining is to interact with the system through the corresponding Postman API collection. This enables the execution of a complete data exchange process between a provider and a consumer using the EDC.

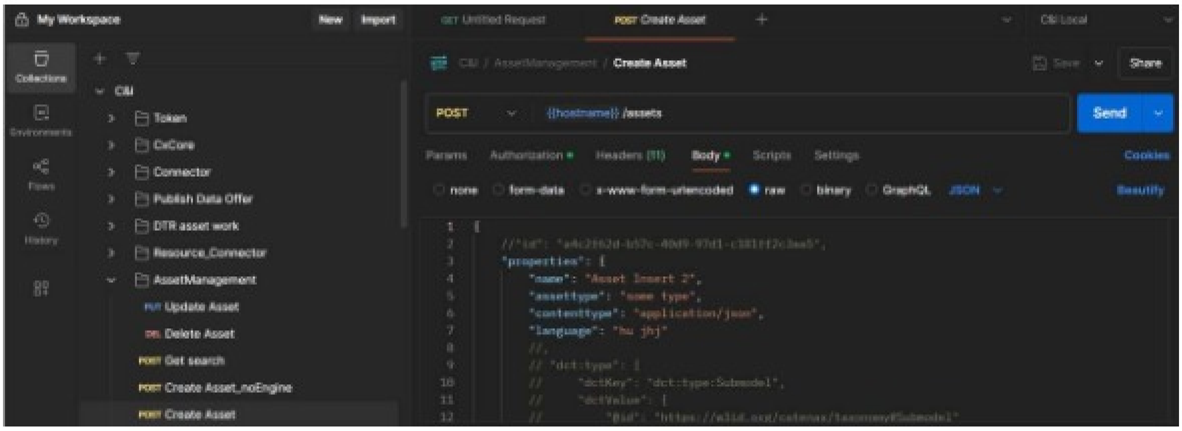


Figure 29. Postman collection to operate the EDC

As shown in Figure 29, the API provided by the connector can be used to create an asset. Similarly, other endpoints allow the creation of a contract definition that links the asset to a specific policy, and the subsequent publication of that asset in the data catalog.

These operations can also be carried out using the graphical interface illustrated in Figure 28. This interface is particularly convenient for uploading simpler data objects—such as Excel or CSV files—directly through the external platform's UI. However, when working with more complex data structures accessible via a URL, it is generally more efficient to use the Postman collection.

Now that the connector management is externalized, establishing a data transfer becomes significantly more straightforward. The steps that outline the process required to initiate a data exchange involve creating a Data Asset, defining a Contract Definition that links the Data Asset to a specific access policy, and publish the Contract Definition to the data provider's catalog. However, it is needed a data consumer to discover the published asset and initiate a contract negotiation by submitting a Contract Request. It is important to note that both data provider and data consumer should be using the same Tractus-X EDC version (the latest version which is usable in Cofinity-X is EDC 0.7.0).

6.4 IMPLEMENTATION – STEPS TO DATA TRANSFERS

On the one hand, the immediate objective is to achieve a data transfer using the EDC as a Service, in order to continue gaining familiarity with EDC-based technologies. To that end, and leveraging the externalized as a service connector, the following steps are required:

- By creating a trigger—most likely using Azure Data Factory—the data that Gestamp intends to provide to a data consumer is first stored in Azure Blob Storage.
- Using the Postman API collection, a data asset is then created. This asset is configured to reference the data stored in the Azure Blob Storage container.
- Once the asset is correctly defined and linked to the data in Blob Storage, the data exchange process is initiated by following the standard EDC flow, as described in Chapter 6.

On the other hand, the long-term goal remains to continue developing Gestamp's internal capabilities in EDC technology. For this reason, it is considered necessary to resume work on deploying a self-managed local connector, which will allow independent operation without relying on third-party providers. In other words, the focus should be place on the local EDC and not in the EDC as a service.

Therefore, a deployment roadmap must be defined to host an EDC instance for Gestamp. However, for such an instance to be moved into a production environment, it must undergo an audit by Deloitte. In the meantime, it can be tested temporarily in the Cofinity-X Beta environment.

Since the local connector cannot be deployed due to the hardware limitations of the current host machine, a virtual machine with sufficient resources must be used. For instance, AWS provides dedicated tutorials for deploying EDC connectors using cloud-based infrastructure [37].

Furthermore, using cloud infrastructure is a more scalable alternative to running Kubernetes locally. Kubernetes is a container orchestration platform designed for automating deployment, scaling, and management of containerized applications [38]. When running locally, its scalability is limited by the available resources (CPU, RAM, disk) of the host machine, which restricts the ability to deploy environments such as the EDC, since it has multiple microservices. In contrast, deploying Kubernetes clusters in the cloud provides the flexibility to allocate dynamic resources, enables high availability, and decouples the infrastructure from individual developer machines. This approach supports better performance and maintainability.

Chapter 7. RESULT ANALYSIS

The objective of deploying an EDC and achieving a functional data exchange has been a step towards a future interconnection model of data spaces based on telecommunication networks. However, given that no inter-data space data transfer was attempted, the outcome of this phase must be limited to an analysis of the EDC deployment attempts.

The deployment of the EDC was attempted through five different methods. The first four approaches aimed to develop an in-house EDC but were unsuccessful because the microservices were not initialized correctly. However, the fifth method, consisting in using an EDC as a service from an external provider, has enabled Gestamp to get integrated into Catena-X.

The first tryout to deploy the Tractus-X EDC was not successful because the repository does not offer a ready-to-use dataspace environment and because of the incompatibility issue regarding the EDC version and Cofinity-X.

The deployment attempt using the Tractus-X MXD Chart proved to be ineffective. The repository was outdated and misaligned with current Tractus-X standards, which led to a non-functional setup and prevented the completion of the data exchange. This resulted in a misallocation of time and resources. Yet, it must be noted that the obsolescence of the repository was not indicated to Gestamp beforehand. It was only through consultation with the Tractus-X Matrix Chat, the official Tractus-X community chat that this situation was confirmed, reflecting the lack of active maintenance of the connector documentation.

The attempt using the Tractus-X Umbrella Chart also failed to achieve an EDC deployment. Although the official documentation specified a minimum requirement of 4 CPU cores and 6 GB of RAM, this proved insufficient in a local environment based on Windows OS. Following consultation with the external company *I.* specialized in data spaces, it was concluded that the failure was primarily due to hardware limitations and the operating system used. It was recommended to use Ubuntu-based virtual machines, which offer better resource management for orchestrating Kubernetes clusters. Ubuntu is often preferred in these scenarios due to its lightweight architecture and greater compatibility with Docker and Kubernetes environments [39]. Additionally, using cloud-based infrastructure such as the one provided by AWS, for example, is a more scalable alternative than running Kubernetes with Docker Desktop locally.

A recurrent issue faced across these attempts is the immaturity of the EDC technology and its ecosystem. Much of the official documentation remains underdeveloped or outdated, and practical support often depends on informal (but official) channels like the Matrix Chat rather than structured guides. This creates a steep learning curve and leads to unnecessary delays and inefficiencies.

Due to the limitations of local deployment, an external EDC as a service was obtained in the Cofinity-X Marketplace. This provided a functional environment for testing data asset creation, contract definition, and catalog publication. However, this approach relies on third-party infrastructure and does not contribute to developing in-house deployment capabilities.

In conclusion, integration into Catena-X was successful, but the goal of implementing a complete data transfer within data space has not yet been achieved. Therefore, the IHC development has not yet begun. However, this remains a real use case for the company, and work will continue under an Agile methodology. Moreover, this first iteration into the connector development has been useful to obtain valuable insights into EDC infrastructure requirements or the need to have version compatibility. Therefore, the current priority to continue with the project is to obtain a cloud virtual machine that meets the necessary specifications to develop an in-house EDC.

Chapter 8. CONCLUSIONS AND FUTURE WORK

Data spaces represent a paradigm shift in the way organizations govern data by standardizing data exchange and promoting interoperability across the entire supply chain. Initiatives such as Catena-X, built on the technical principles of Gaia-X and the governance principles of the IDSA, have the potential to revolutionize industrial collaboration and are positioned to transform industries like the automotive one.

Although the original direction of this project was aimed at exploring interconnectivity between data spaces, the implementation challenges encountered led to a pivot towards a more foundational objective: the integration and deployment of the EDC into Cofinity-X. In other words, the focus shifted to the preliminary step of making an intra-data space transfer rather than to focus on the next step, which would be transferring data between data spaces. This redirection aligns with the principles of the Lean Startup methodology [40] where iterative development allow projects to evolve in response to constraints and feedback from the environment.

Throughout the course of this project, the immaturity of the EDC ecosystem becomes evident. Despite the architecture of the EDC seeming to be well-established, sveeral obstacles arose due to outdated documentation, which prevented the full deployment of a self-managed EDC instance. As a consequence, the prototyping of an IHC and submitting the architecture proposal to Gaia-X for potential adoption was not started.

It is important to highlight that this initiative reflects a real and dynamic use case in Gestamp, and therefore, it is affected by the priorities of the company and, since the beginning of the project up to the current date, internal priorities have shifted. The urgency of interconnecting multiple data spaces has been lowered, while the necessity of establishing a reliable connection to an existing data space—and performing validated data transfers—has gained importance.

Although the goal of demonstrating the Digital Product Passport (DPP) exchange was not reached, this remains a key milestone for future iterations. The DPP is a critical component in future industrial data models, especially for regulatory compliance and supply chain transparency.

Regarding future work, the following actions are proposed to continue the development:

1. Achieve data transfer in Catena-X using the provided EDC as a service.

2. Achieve data transfer in Catena-X using a self-managed EDC instance.

3. Implement basic interconnectivity between two data spaces. Test bidirectional communication between connectors to simulate a data network.

4. Deploy and validate a DPP use case. Once the EDC is operational, the DPP model can be used to test data transfer with product information across connectors.

5. Integrate PCF functionalities. Beyond DPP, extending the EDC to support PCF data exchange adds significant value. PCF data spaces facilitate the sharing of lifecycle emissions data, enabling accurate PCF calculations across organizational boundaries [41] - these data flows are crucial for sustainability reporting and alignment with environmental standards.

In the medium to long term, PCF data exchange could become a strategic pillar for companies like Gestamp. It supports alignment with key ESG objectives, and is central to fulfilling global and regional frameworks such as the Paris Agreement [42], the UN Net Zero 2050 roadmap [43], and the EU climate targets for reducing Scope 1, 2, and 3 emissions[6].

These sustainability objectives are increasingly integrated into corporate KPIs and reporting frameworks [44]. By making a combination of ESG models like PCF and DPP with data

[6] Scope 1: Emissions produced by the direct activity of the company | Scope 2: Emissions produced from the purchase of electricity needed to perform the direct activity form the company | Scope 3: Emissions produced directly by producers and clients in the supply chain of the company.

space technological development, organizations create long-term strategic value. Using technology to find sustainable solutions represents placing engineering at the service of society, being competitive at an industrial level and environmentally responsible.

Lastly, it is worth noting that this work can be linked to the Bachelor's Thesis in Business Analytics titled "Predictive model of CO_2 emissions in the supply chain of Gestamp, an automotive sector company", written by myself, Ignacio Núñez Gómez, and supervised by Atilando Ramiro Fernández-Pacheco Sánchez Migallón. This thesis involves a prediction of Gestamp's Scope 3 emissions resulting from the purchase of steel for the manufacturing of automotive parts. Both projects—the Bachelor's Thesis and this Master's Thesis—could generate synergies, as the company's emissions forecast could be shared with suppliers and customers in the automotive value chain through Catena-X.

GLOSSARY OF ACRONYMS

- **AISBL:** Association Internationale Sans But Lucratif
- **API:** Application Programming Interface
- **AP:** Advisory Provider
- **BAP:** Business Application Provider
- **BPN:** Business Partner Number
- **CAB:** Conformity Assessment Body
- **CPU:** Central Processing Unit
- **CSP:** Core Service Provider
- **CSV:** Comma-Separated Values
- **DHT:** Distributed Hash Table
- **DID:** Decentralized Identifier
- **DNS:** Domain Name System
- **DNSSEC:** Domain Name System Security Extensions
- **DPC:** Data Providers and Consumers
- **DPP:** Digital Product Passport
- **DSP:** Data Space Protocol
- **EDC:** Eclipse Dataspace Connector
- **eIDAS:** electronic IDentification, Authentication and trust Services
- **ESG:** Environmental, Social and Governance
- **ESP:** Enablement Service Provider
- **FTP:** File Transfer Protocol
- **GBA:** Global Battery Alliance
- **GUID:** Globally Unique Identifier
- **GXDCH:** Gaia-X Digital Clearing House
- **HCL:** Hashicorp Configuration Language
- **HTTP:** HyperText Transfer Protocol

- **IaC :** Infrastructure as Code
- **IAM:** Identity and Access Management
- **ICAM:** Identity, Credential and Access Management
- **IDSA**: International Data Spaces Association
- **IDTA:** Industrial Digital Twin Association
- **IH:** Interconnectivity Hub
- **IHC:** Interconnectivity Hub Connector
- **IP:** Internet Protocol
- **ISP:** Internet Service Provider
- **K8s:** Kubernetes
- **KIT:** Keep It Together
- **MAC:** Media Access Control
- **MXD:** Minimum Tractus-X Dataspace
- **OS:** Operating System
- **OSP:** Onboarding Service Provider
- **OSI:** Open Systems Interconnection
- **PCF:** Product Carbon Footprint
- **QoS:** Quality of Service
- **RAM:** Random Access Memory
- **R&D:** Research and Development
- **SMTP:** Simple Mail Transfer Protocol
- **SPI:** Service Provider Interface
- **SSI:** Self-Sovereign Identity
- **SSL:** Secure Sockets Layer
- **TCP:** Transmission Control Protocol
- **TLS:** Transport Layer Security
- **UDP:** User Datagram Protocol
- **UI:** User Interface
- **URL:** Uniform Resource Locator

- **VC:** Verifiable Credential
- **vCPU:** Virtual Central Processing Unit
- **WBCSD:** World Business Council for Sustainable Development
- **YAML:** Yet Another Markup Language

BIBLIOGRAPHY

[1] [En línea]. Available: https://www.cofinity-x.com/.

[2] «Eclipse Tractus-X,» Eclipse, 2024. [En línea]. Available: https://eclipse-tractusx.github.io/.

[3] Docker, «Use containers to Build, Share and Run your applications,» docker.com, [En línea]. Available: https://www.docker.com/resources/what-container/#:~:text=A%20Docker%20container%20image%20is,tools%2C%20system%20libraries%20and%20settings..

[4] Centro de Desasrrollo de Competencias Digitales Castilla-La-Mancha, «Arquitectura de software: microservicios, contenedores y orquestación,» bilib.es, 06 February 2019. [En línea]. Available: https://www.bilib.es/actualidad/articulos-tecnologicos/post/noticia/arquitectura-de-software-microservicios-contenedores-y-orquestacion.

[5] kubernetes.io, [En línea]. Available: https://kubernetes.io/.

[6] M. Aleksic, «Kubernetes Pods: Basics for Beginners,» phoenixnap.com, 9 March 2023. [En línea]. Available: https://phoenixnap.com/kb/kubernetes-pod.

[7] A. Skotnický, «What is Kubernetes and Why is it the Future of Cloud Computing,» taikun.cloud, [En línea]. Available: https://taikun.cloud/what-is-kubernetes-and-why-is-it-the-future-of-cloud-computing/.

[8] AWS, «¿Cuál es la diferencia entre Terraform y Kubernetes?,» aws.amazon.com, [En línea]. Available: https://aws.amazon.com/es/compare/the-difference-between-terraform-and-kubernetes/?utm_source=chatgpt.com.

[9] "SUPPLYON," [Online]. Available: https://www.supplyon.com/en/industries/automotive_industry/catena-x/.

[10] «DLR - Institute for AI Safety and Security,» [En línea]. Available: https://www.dlr.de/en/ki/research-transfer/projects/catena-x#:~:text=The%20Catena%2DX%20Automotive%20Network,and%20on%20to%20recycling%20businesses.

[11] «Gestamp se une a Catena-X para fomentar la colaboración en la industria de la automoción,» Gestamp, 2 October 2023. [En línea]. Available: https://www.gestamp.com/Medios/Prensa/Comunicados-de-prensa/2023/Gestamp-se-une-a-Catena-X-para-fomentar-la-colaboracion-en-la-industria-de-la-automocion.

[12] O. Hölck, S. Voges, K.-F. Becker, E. Wagner y M. Schneider-Ramelow, «Why Should We Share Material Data? The Positive Impact of Public/Private Databases Access Through Digital Twins Along the Product Value Chain,» *IEEE,* 2024 June 2024.

[13] «GAIA-X Spain,» GAIA-X, [En línea]. Available: https://www.gaiax-spain.com/.

[14] «Catena-X: Network for Cross-Company Data Exchange in the Automotive Industry Relies on IDS,» International Data Spaces Association, 3 March 2021. [En línea]. Available: https://internationaldataspaces.org/catena-x-network-for-cross-company-data-exchange-in-the-automotive-industry-relies-on-ids/.

[15] F. Schöppenthau, F. Patzer, B. Schnebel, K. Watson, N. Baryschnikov, B. Obst, Y. Chauhan, D. Kaever, T. Usländer y P. Kulkarni, «Building a Digital Manufacturing as a Service Ecosystem,» *MDPI,* 2023.

[16] P. Leschinski, «Catena-x,» 2023. [En línea]. Available: https://catena-x.net/de/angebote-standards/hmi-2023/circular-economy-1.

[17] E. Union, «European Data,» European Union, 27 September 2024. [En línea]. Available: https://data.europa.eu/en/news-events/news/eus-digital-product-passport-advancing-transparency-and-sustainability.

[18] Agile, [En línea]. Available: https://agilemanifesto.org/.

[19] Gaia-X, «Gaia-X Architecture Document,» 2024.

[20] Gaia-X, «Association,» Gaia-X, [En línea]. Available: https://gaia-x.eu/who-we-are/association/.

[21] Gaia-X, «Trust framework - Gaia-X Trust Framework - 22.04 Release,» 2022.

[22] Gaia-X, «Gaia - X Compliance Document,» 2024.

[23] Gaia-X, «Identity, Credential and Access Mnagement Document,» 2024.

[24] Auth0, «What is OpenID Connect (OIDC)?,» Auth0, [En línea]. Available: https://auth0.com/intro-to-iam/what-is-openid-connect-oidc.

[25] Catena-X, «Catena-X Operating Model,» 2023.

[26] «Eclipse Dataspace Components,» Eclipse Foundation, [En línea]. Available: https://projects.eclipse.org/projects/technology.edc.

[27] «Gaia-X Digital Clearing House (GXDCH),» Gaia-X, [En línea]. Available: https://gaia-x.eu/services-deliverables/digital-clearing-house/.

[28] European Union, «Discover eIDAS,» European Comission, [En línea]. Available: https://digital-strategy.ec.europa.eu/en/policies/discover-eidas.

[29] «EONA-X,» LinkedIn, [En línea]. Available: https://www.linkedin.com/company/eona-x/.

[30] «EONA-X Data Space,» VivaTechnology, [En línea]. Available: https://vivatechnology.com/partners/eona-x-data-space.

[31] International Data Spaces Association (IDSA), «Data Spaces Business Models,» 2024.

[32] Cloudflare, «DNS root server,» cloudflare.com. [En línea].

[33] Hazelcast, «What Is a Distributed Hash Table?,» [En línea]. Available: https://hazelcast.com/foundations/distributed-computing/distributed-hash-table/.

[34] IBM, «¿Qué es DNSSEC (extensiones de seguridad DNS)?,» ibm.com, [En línea]. Available: https://www.ibm.com/es-es/topics/dnssec#:~:text=Khan%2C%20Michael%20Goodwin-,%C2%BFQu%C3%A9%20es%20DNSSEC%20(extensiones%20de%20seguridad%20DNS)?,se%20alteran%20en%20el%20camino..

[35] Catena-X, catena-x.net, [En línea]. Available: https://catena-x.net/.

[36] Think-it Team, «The EDC Connector: Driving Data Sovereignty,» Think-it, 4 November 2024. [En línea]. Available: https://think-it.io/insights/edc-connector.

[37] I. A. (.-i. M. G. (.-i. J. H. S. (. a. M. M. (. Created by Ramy Hcini (Think-it), «Set up a minimum viable data space to share data between organizations,» AWS, [En línea]. Available: https://docs.aws.amazon.com/prescriptive-guidance/latest/patterns/minimum-viable-data-space-share-data-organizations.html.

[38] S. Caron, «What is Kubernetes? K8s explained,» dynatrace, 8 April 2025. [En línea]. Available: https://www.dynatrace.com/news/blog/what-is-kubernetes-2/#:~:text=Kubernetes%20is%20a%20popular%20solution,and%20microservices%20than%20ever%20before..

[39] ComputerWorld, «Seis distribuciones de Kubernetes que lideran la revolución de los contenedores,» computerworld.es, [En línea]. Available: https://www.computerworld.es/article/2112884/seis-distribuciones-de-kubernetes-que-lideran-la-revolucion-de-los-contenedores.html#:~:text=Uno%20de%20los%20grandes%20puntos,impulsadas%20por%20CPU%20y%20GPU..

[40] E. Ries, The Lean Startup, 2011.

[41] T. Hahn, «Can data spaces support sustainability in manufacturing?,» europeanfiles.com, 7 February 2023. [En línea]. Available: https://www.europeanfiles.eu/digital/can-data-spaces-support-sustainability-in-manufacturing#:~:text=Data%20spaces%20can%20enable%20cross%2Dcompany%20sharing%20of,carbon%20footprint%20(PCF)%20by%20companies%20along%20the.

[42] UNFCCC, «The Paris Agreement,» unfccc.int, [En línea]. Available: https://unfccc.int/process-and-meetings/the-paris-agreement#:~:text=What%20is%20the%20Paris%20Agreement?&text=The%20Paris%20Agreement%20is%20a,and%20adapt%20to%20its%20effects..

[43] IEA, «Net Zero by 2050,» iea.org, [En línea]. Available: https://www.iea.org/reports/net-zero-by-2050.

[44] European Commission, «Guidelines on reporting climate-related information,» Bruxelles/Brussel, 2019.

[45] United Nations, «THE 17 GOALS,» un.org, [En línea]. Available: https://sdgs.un.org/goals.

[46] «Wikipedia,» [En línea]. Available: https://en.wikipedia.org/wiki/Gaia-X.

[47] Hub France Gaia-X, «YouTube,» [En línea]. Available: https://www.youtube.com/watch?v=wkBaAK0y7xA&t=71s.

[48] HashiCorp, «What is Vault?,» developer.hashicorp.com, [En línea]. Available: https://developer.hashicorp.com/vault/docs/about-vault/what-is-vault.

[49] Microsoft, «Identity Hubs as Personal Datastores,» techcommunity.microsoft.com, [En línea]. Available: https://techcommunity.microsoft.com/blog/microsoft-entra-blog/identity-hubs-as-personal-datastores/389577.

[50] J. Schmitt, «circle ci blog,» circleci.com, 6 January 2025. [En línea]. Available: https://circleci.com/blog/what-is-helm/#:~:text=Helm%20charts,an%20application%20with%20different%20configurations..

ANNEX: PROJECT ALIGNMENT WITH THE SDGS

This project is closely aligned with several Sustainable Development Goals [45]. Some of them are described in this Annex.

SDG 9: INDUSTRY, INNOVATION AND INFRASTRUCTURE

The goal of this SDG is to "Build resilient infrastructure, promote inclusive and sustainable industrialization, and foster innovation." This aligns perfectly with the core of data space technology: creating robust, reliable, and progressive data exchange infrastructures.

Data spaces are a disruptive and cutting-edge technology, growing at an accelerated pace. This project revolves around this very SDG—innovating in data exchange technology, with a strong emphasis on European technologies, and focusing specifically on innovating and optimizing within supply chains across various industries, such as the automotive sector.

Therefore, it can be concluded that SDG 9 is a key priority for the scope of this project.

SDG 12: RESPONSIBLE CONSUMPTION AND PRODUCTION

Data spaces promote consumption and, especially, responsible production, by enabling greater transparency and traceability across the product lifecycle. Through initiatives like the Digital Product Passport (DPP), companies can access key information about how a product was made, used, and recycled.

This visibility helps businesses optimize resource use, reduce waste, and design more sustainable products, since they have more information about what phases in the process of developing a product can be improved. If there is information about how a product is created, there is space for improving the process. This way, data spaces contribute directly to SDG 12, fostering sustainability in industrial ecosystems.

SDG 13: CLIMATE ACTION

One of the intrinsic goals of data spaces is their contribution to climate change mitigation. A key use case of initiatives like Catena-X is the implementation of the Product Carbon Footprint (PCF) mechanism. This enables companies to exchange reliable and standardized data on CO_2 emissions associated with products across the entire supply chain.

To portray this, an example will be described: if a company like Gestamp can access accurate and verifiable emissions data from two potential suppliers via the Catena-X data space, it can select the one with a lower carbon footprint. This allows Gestamp to reduce its Scope 3 emissions, effectively lowering its overall environmental impact, and increasing competitiveness between suppliers, pushing them to reduce their emissions. Moreover, if Gestamp has access not only to suppliers from its own data space, but it can also receive data from suppliers in other data spaces (as proposed in this project), competitiveness to reduce emissions between suppliers is boosted.

By enabling mechanisms like the PCF of the DPP, data spaces like Catena-X actively support SDG 13 – Climate Action, placing technology and data exchange at the service of global decarbonization efforts.

SDG 17: PARTNERSHIPS FOR THE GOALS

At their core, data spaces are collaborative ecosystems designed to enable secure data exchange between organizations. These environments are fundamentally built on trust to build partnerships and strategic alliances across companies.

A concrete example is Catena-X: its ultimate goal is to drive innovation in the industry by enabling more sustainable production models, improving resource efficiency, and supporting climate action—all through data sharing and within a common framework of partnerships among manufacturers, suppliers, and service providers. If Catena-X is able to communicate with other data spaces as proposed in this project, a global network of partnerships could be established.

By promoting interoperability, trust and collaboration, data spaces contribute directly to SDG 17, strengthening partnerships that are critical for achieving the broader 2030 Agenda goals.