INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO
MESTRADO EM ENGENHARIA INFORMÁTICA




A Framework Gaia-X e sistemas IOT para aplicação em contexto do Sector da Energia

RICARDO MIGUEL DE CARVALHO BARBOSA BARROS
Setembro de 2024

POLITÉCNICO
DO PORTO

ISEP INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO

The Gaia-X Framework and IoT systems for application in the context of the energy sector

Ricardo Miguel de Carvalho Barbosa Barros

Dissertation Plan for the Degree of master's in Computer Engineering, Specialization in Software Engineering

Supervisor: Luís Miguel Pinho

Co-Supervisor: Carlos Adriano Gonçalves

Porto, September 2024

Statement of Integrity

I hereby declare having conducted this academic work with integrity.

I have not plagiarized or applied any form of undue use of information or falsification of results along the process leading to its elaboration.

Therefore, the work presented in this document is original and authored by me, having not previously been used for any other end.

I further declare that I have fully acknowledged the Code of Ethical Conduct of P.PORTO.

ISEP, Porto, September 15, 2024

Acknowledgements

I would like to thank my advisors Luis Pinho and Carlos Gonçalves for the opportunity and trust to work on this dissertation and for the available help and patience.

I would like to thank my parents Silvano Barros and Ana Paula Barros for the their support through my academic journey and support.

I would also like to extend a special word of thanks to my grandmother Lurdes Spencer, my sister Débora Barros for believing in me and providing me with all the motivation necessary to complete this dissertation.

I would also like to thank my friends, mainly Débora Rocha, for their years of support, valuable advice and patience.

Thank you all.

Resumo

O Gaia-X visa desenvolver uma infraestrutura de dados comum, segura e soberana para a Europa, que permita ao proprietário dos dados manter o controlo total e a soberania sobre os dados, enquanto promove a colaboração entre diferentes sectores. Esta tese irá, por conseguinte, realizar investigação sobre a aplicação do Gaia-X ao sector da energia, especificamente no que diz respeito ao aproveitamento da partilha descentralizada de dados para uma melhor gestão e inovação da energia. Com o Gaia-X, foi planeado para ser segura, transparente e eficiente, enquanto orienta a transição para sistemas de energia mais sustentáveis.

A tese analisa os *data spaces*, desde a sua definição e funcionamento como também o ecossistema, a arquitetura da Gaia-X, desde os serviços de federação, passando pelos mecanismos de conformidade, até aos componentes necessários para implementar um *data space* seguro. O protótipo foi desenvolvido sobre uma base do Gaia-X Digital Clearing House. Esta base ainda se encontra numa fase inicial de desenvolvimento, sendo por isso ainda instável e com documentação pormenorizada escassa, no entanto constitui uma base de trabalho para demonstração do conceito.

Apesar destas limitações, a tese apresenta informações pertinentes sobre os desafios e oportunidades que acompanham a implantação do Gaia-X no sector da energia. Salienta a necessidade de um maior desenvolvimento, tanto a nível das normas técnicas para uma adoção em grande escala. A tese corrobora o potencial do Gaia-X para apoiar ecossistemas de dados seguros e descentralizados, enquanto fator essencial para uma gestão inovadora e sustentável da energia, sugerindo assim direções para mais investigação e aplicações práticas.

Palavras-chave: Gaia-X, IoT, interoperabilidade, soberania de dados, sector de energia, data spaces, serviços de federação, descentralizada.

Abstract

Gaia-X aims to develop a common, secure, and sovereign data infrastructure for Europe that will enable the data owner to retain full control and sovereignty over the data while fostering collaboration between different sectors. This dissertation will therefore be conducting research on the application of the Gaia-X framework to the energy sector, specifically in leveraging decentralized data sharing for better management and innovation of energy. With Gaia-X has been planned to be secure, transparent, and efficient while guiding the transition towards more sustainable energy systems.

The dissertation analyzes data spaces, its inner working and its ecosystem, the architecture of Gaia-X, from federation services over compliance mechanisms to components needed to implement a secure dataspace. Considering that the prototype was developed with many limitations, such as the early stage in development of the Gaia-X Digital Clearing House, and given that detailed documentation is scarce, it forms a very good basis with which one can improve.

Despite these limitations, the dissertation provides valuable insights into challenges and opportunities that accompany the deployment of the Gaia-X framework in the energy sector. It highlights the necessity for further development both at the level of technical standards and at that of regulatory frameworks for large-scale adoption. Nevertheless, this dissertation corroborates the potential of the Gaia-X framework to support secure, decentralized data ecosystems as the key enabler for innovative and sustainable energy management, thus suggesting directions for further research and practical applications.

Keywords: Gaia-X, IoT, interoperability, data sovereignty, energy sector, data spaces, federation services, decentralized.

Index

Figures Index

Tables Index

Acronyms

List of Acronyms

IoT – Internet of Things

IDS – International Data Spaces

ICT – Information and Communication Technology

GXDCH – Gaia-X Digital Clearing House

VC – Verifiable Credentials

SSI – Self-Sovereign Identity

ICP – Interconnection Point Identifier

API – Application Programming Interface

PSD2 – Payment Services Directive

GDPR – General Data Protection Regulation

EIDAS – Electronic Identification, Authentication, and trust Services

EBSI – European Blockchain Service Infrastructure

1 Introduction

1.1 Context

The concept of data space provides a very real paradigm shift in sharing, managing, and using data across Europe. These secure and interoperable environments provide participants with a way to share data while retaining control or, if you will, sovereignty over their information. In contrast with other methods of sharing data, data spaces enable data to remain decentralized-or, in other terms, resident at the source-even while it is shared out for integration and analytics based on standardized protocols and interfaces. This approach enhances data security, reduces duplication, and supports all kinds of data formats for innovation and scalability (Team Data Spaces, 2024).

To put this into context, Gaia-X is an initiative within the general European strategy of establishing a data infrastructure that is federated and decentralized-one that will offer assurances of transparency, trust, and much-needed sovereignty in this digital space underpinning European values. Driven by the concept of cultivating a collaborative digital economy, Gaia-X pursues a goal for innovation at the heart of traditional industries such as energy, health, and manufacturing sectors (Gaia-X Association, 2024a).

It is under these premises that this dissertation finds its context in the Gaia-X initiative, with a narrow focus on its application in the energy sector. The key objective of this work will be to investigate and analyze the architecture and operation principles of the Gaia-X, especially those that allow for realizing at least a minimum viable energy dataspace. This will be done by explaining the main components of Gaia-X-like federation services and compliance mechanisms-and showing their applicability within a real energy management scenario.

Work presented here should contribute to the ever-growing knowledge on Gaia-X with a practical case that will meet its standards and principles. To this end, this should serve to illustrate the potential of how Gaia-X can enable secure, interoperable, and efficient data sharing for the energy sector in fostering more sustainable and effective energy management practices across Europe.

1.2 Objectives

The purpose of this thesis is to analyze the framework for Gaia-X, focusing on its basic architecture and the principles behind its mode of operation. In other words, its ultimate objective is a detailed understanding of Gaia-X's structure, the evaluation of its effect on data infrastructure in various sectors, and research on how far it could contribute to enhanced data sovereignty, security, and interoperability. It will thus follow the following concrete objectives:

- **Analysis of the Gaia-X framework:** Accordingly, make a profound analysis of the architecture of Gaia-X identify and document its key components, and investigate how these components are interlinked in such a way that it will be cohesive and functional.

- **Design a Prototype System with the Gaia-X framework**: A prototype system shall be designed and developed based on the directives from Gaia-X, specifically including the data space concept in a way that it can best illustrate practical applicability, especially to energy system management and optimization.

- **Demonstrate the Utilization of Gaia-X in Energy System Management**: The prototype will be used in the real world in the energy sector. This will show how Gaia-X framework will enhance data sharing, security, and interoperability within the current trends in energy management.

1.3 Methodology Approach

The chosen methodology for this dissertation is composed of the following six guidelines:

1. Literature review – The most important, it involves thorough review of existing literature for a better understanding of current knowledge.

2. Case study scenario – Involves selection and preparation of case study relevant to answer the research questions.

3. Development of the conceptual model and architecture – Involves theorizing a model / architecture based on the literature review and case study.

4. Implementation and testing – Critical step for testing the validity and reliability of our model.

5. Analysis of result: Consists of analyzing the collected data from testing to make a thorough analysis in hopes of answering the research question.

6. Dissertation writing and documentation – Compiling the research findings, analysis, and discussions.

Each of the guidelines is fundamental to a rigorous research process and contribution to the dissertation's quality and integrity.

1.4 Structure and Organization

The present document is organized in six sections:

1. **Introduction –** This section provides an overview of the status in data infrastructures and the need for such a decentralized, secure, interoperable framework like Gaia-X, particularly for those sectors which demand high-level data sovereignty and security.

2. **Literature Review –** This section addresses the definitions, challenges regarding the implementation of data spaces in Europe, and design and architecture principles. It reflects on very basic constituent elements of data spaces, like decentralized data, interoperability, and data sovereignty, and then develops some of the hardest issues- which are related to governance, trust management, and regulatory compliance. It also covers how Self-Sovereign Identity can be applied to reinforce privacy and user control, provides a comparison between the International Data Spaces

3. **Gaia-X Ecosystem –** This section talks about the architecture of the Gaia-X and the operational model. It explains how Gaia-X creates a connected, federated, secure data environment, connecting several cloud service providers while enforcing data sovereignty and security.

4. **Realizing Energy Dataspace** - This Section discusses a practical use case: the implementation of Gaia-X within the energy sector. It goes through steps and technologies necessary to realize a compliant energy dataspace, using Gaia-X Digital Clearing House and Dataspace Connector provided by Sovity. It describes how the data sharing workflow would include activities such as setting policies, service contracts, and compliance checks for secure and efficient data exchange.

5. **Critical Analysis** - The Critical Analysis chapter evaluates how successful the framework of Gaia-X is in achieving its objectives. It provides a critical analysis of the strengths and weaknesses of Gaia-X. The chapter also discusses some practical implementation challenges that have arisen.

6. **Conclusion** – This section wraps up with the key findings of the research study, reflecting on the implication of real-world applications that could be carved out using Gaia-X. The main discussion here pertains to the probable effect Gaia-X would have on

increasing the rate of data sharing with security and compliance in different sectors, the most important of which is the energy sector, considering it as one of the highly regulated sectors. It also concludes with limitations found while researching the study and suggests areas of further study, considering how the digital world, along with its regulatory framework, is under continuous evolution.

2 Literature Review

2.1 Introduction

Data exchange traditionally deals with the transfer of data between different parties for some limited usage internally, data sharing works around dynamic and collaborative engagement in view of fostering innovation and efficiency across whole ecosystems. While data exchange is bound by its scope and purpose, data sharing means open and flexible use for collective problem-solving and value creation. To have an effective data sharing it requires a well-organized environment, appropriate governance framework, and explicitly defined principles of data sovereignty (Boris and Hompel, 2022).

Boris and Hompel (2022) discusses that the European data spaces are a key concept for the future, and one such example is Gaia-X, a secure, federated data infrastructure that includes European values such as privacy, security, and data sovereignty. By providing a standardized framework for sharing data across industries, simultaneously, Gaia-X supports currently faced issues of data sharing but also sets up grounds for a more integrated and competitive digital economy in Europe. The initiative is likely to change the playing field of innovation driven by data and ensure that data sharing remains a powerful means of driving collaboration and efficiency in a connected world.

2.2 What is a data space

In simple terms a data space can be defined as secure and standardized digital infrastructure that enables trusted data exchange and data-base services (International Data Space Association (IDSA), 2024). But that's only one definition as we have several entities such as the International Data Space Association and Gaia-X Association that have their own definition of what a data space is.

The International Data Space Association (2024), for example defines a data space as a virtual space that provides a standardized framework for data exchange, based on common protocols and formats, as well as secure and trusted data sharing mechanisms. Meanwhile the Gaia-X Association (2024a), defines data space as a type of data relationship between trusted partners who adhere to the same high-level standards and guidelines in relation to data storage and sharing withing one or many specific industries.

Each data space involves the participants being data providers, users or data consumers, and intermediaries (Gaia-X Hub, 2022). One can argue that perhaps some of the most essential components of data spaces are data sharing and data sovereignty. Data sharing can be described as the exchange of data or even the deployment of data processing capabilities among different involved parties and data sovereignty simply meaning putting stakeholders

back in control regarding their personal data, such as management of digital processes, infrastructures, and data flows for which appropriate governance must be put in place. These data spaces are primarily tailored to realize secure data exchange between participants. There is a whole group of subfunctions backing this up, which administer and control the provisioning, receipt, and transmission of data. For a data space to be effective, following attributes must be present such as **scope** (a data space should define its scope by specifying the data type, participants, and duration), **decentralization** (an effective data space is decentralized and reduces reliance on any single entity), **federation or interoperability** (it eases data flow between diverse participants and spaces), **transparency** (participants must be clearly informed about data exchanges in the data space), **sovereignty** (participants should make all decisions about participation and usage) and finally **trust** (transparency and compliance with legal standards build trust among participants). Each of these helps ensure that a data space is to be robust, secure and aligned with the values set by Europe (Reiberg et al., 2022).

Understanding what a data space is, and its most essential attributes are key to understanding what Gaia-X offers and how it can be a betterment for data sharing and data sovereignty in general.

2.3 Different perspectives of data spaces

Data spaces are complex in nature and can be looked from three main perspectives: technical, economic, and legal. Each of these views adds up to provide various insights into the development and putting into practice of data spaces in such a manner that it transforms data exchange and use in sectors (Reiberg et al., 2022).

From **technical perspective** data spaces are solutions for data integration that enable access to and the use of data that is distributed without having to move it in its entirety and permanently. In their federated architecture, data can be located at source but is still accessible for integration and analysis. This enhances security and reduces duplications. They support a large variety of formats and sources of data through the adoption of standardized protocols and interfaces, thus stimulating innovation and scalability when new requirements or technologies need to be faced.

From **economic perspective** data spaces are of high economic value because they realize digital economic exchange and innovation by means of digital data flows between companies, entrepreneurs, and consumers, ultimately leading to new business models and markets. They also contribute to economic and geopolitical sovereignty by placing power in the hands of any organization to use data for competitive advantage and growth. Incentives from an economic point of view to take part in data spaces create enhanced efficiency, improved customer experience, and new revenue streams.

From **legal perspective** data spaces offer a legal framework under which data would be controlled and used according to the wishes and conditions of the parties concerned. This is

important now when data privacy and protection are issues of paramount concern. Legal frameworks, such as the General Data Protection Regulation and the Data Governance Act, ensure harmonized standards and requirements for intermediaries in establishing safeguards to protect users' rights while ensuring that data exchanges are conducted in concurrence with national and international laws. In this way, data spaces alleviate legal barriers even more by promoting neutrality and accountability for data intermediaries through the installation of governance and oversight mechanisms, ensuring that data exchange occurs fairly, transparently, and in accordance with the law and social expectations.

Taking into consideration all the perspectives we talked about we can see a common denominator in the form of sovereignty. To have sovereignty the participants should make all the decisions about the participation and usage but to achieve this we need the support of regulations such as GDPR, EIDAS etc. for data spaces.

2.4 Supportive Regulations and Recommendations for Data Spaces

Data spaces can be established and start operating under very helpful regulations and recommendations. In that respect, certain frameworks ensure efficient and secure data sharing, and that data exchange occurs in line with applicable laws while increasing the level of trust among stakeholders. The principal regulations and recommendations setting the foundation for the creation of European data spaces have basically been founded on the principles of FAIR—Findability, Accessibility, Interoperability, and Reusability—and entail a few EU regulations and directives (Ulrich Ahle, 2021).

We are going to address the following:

- General Data Protection Regulation (GDPR)
- Electronic Identification, Authentication and trust Services (EIDAS)
- Payment Service Directive (PSD2)
- Context Broker: CEF Building Block
- The European Blockchain Service Infrastructure (EBSI)

General Data Protection Regulation (GDPR) - The GDPR is an all-encompassing data privacy and security law in the European Union, enacted in 2018. It applies to all organizations around the globe that participate in processing personal data belonging to residents of the EU. The law is designed to protect the privacy of individuals by putting them in control of their data while forcing companies to observe strict standards for data protection. Penalties for non-compliance are heavy. Transparency, security, and accountability form the core of GDPR in handling personal data (Wolford, 2024). Ulrich Ahle (2021),explains that in the context of data spaces, each stakeholder, including data consumers, providers, application providers, and capability providers, will have to commit themselves to the protection of individuals' privacy and security of data under GDPR.

Electronic Identification Authentication and Trust Services (EIDAS) - The EU's Electronic Identification, Authentication and Trust Services regulation aims to create trust for electronic transactions between member states. This regulation establishes a legal framework for electronic identification and trust services in the use of electronic signatures, seals, timestamps, and others, ensuring cross-border recognition. Cross-border digital services under this regulation are secured, and it enables persons, businesses, and public authorities to have easy access to services all over Europe (European Commission, 2024). In data spaces, eIDAS is basic for checking digital identity and securing trust in data transactions (Ulrich Ahle, 2021).

Payment Service Directive (PSD2) should contribute to the creation of a more integrated European payment market, raise consumer protection, promote innovation, and improve the security of internet payments and access to accounts within the EU. The directive could also extend to data spaces involving financial transactions or any kind of data monetization, thus opening ways of safe payment procedures with consumer data protection (Ulrich Ahle, 2021).

The Context Broker, proposed by the European Commission under the Connecting Europe Facility initiative, allows for the integration and sharing of pooled data across different sectors. It is useful in data spaces for managing and exchanging real-time data and digital attributes between several stakeholders (Ulrich Ahle, 2021).

The European Blockchain Service Infrastructure (EBSI) functions within the framework of the CEF program, establishing a network of decentralized nodes throughout Europe to facilitate cross-border services related to blockchain and distributed ledger technology. It fosters reliable data transactions and has the potential to improve both the security and traceability of data exchanges within data spaces. Now that we have a good grasp of the available options to guarantee efficient and secure data sharing and ensure data sovereignty, we must be aware of the challenges presented when creating a data space in Europe (Ulrich Ahle, 2021).

2.5 Self-Sovereign Identity (SSI)

To ensure that a user retains its power again with the possibility of regaining control over his own identity and data, Gaia-X resorts to the usage of Self-Sovereign Identity or SSI for short.

Being a fast-growing area, SSI is attracting much interest in academia and industry due to its potential to change identity management in many domains, such as finance, healthcare, and the Internet of Things. It holds in its future the promise of giving users a better grip on their digital lives and, with this, fostering higher levels of trust and security in digital interactions. SSI is a major change in the management of digital identity, shifting centralized control and ownership of personal identity data to the individual. Unlike earlier systems that are based on a central authority managing and validating user information, SSI will decentralize this technology to offer users self-sovereignty over their identity (Š. Čučko and M. Turkanović, 2021).

The core principles of SSI include decentralization, user autonomy, and enhanced privacy and security. By eliminating the need for centralized intermediaries, SSI reduces the risk of data breaches and unauthorized data sharing, which are prevalent in conventional identity management systems. This user-centric approach ensures that individuals have full control over their digital identities, including who has access to their personal data and under what conditions.

Technologies, such as blockchain and Decentralized Identifiers, make self-sovereign identity a reality by offering a secure and verifiable mechanism for how individuals can independently authenticate their identity without any central authority. Verifiable Credentials allow a person to share only the relevant information needed for any given transaction, enhancing privacy and

greatly reducing overexposure to data. In doing so, it reduces many privacy problems directly related to conventional frameworks of identity management.

The image below represents the evolution of identity management.

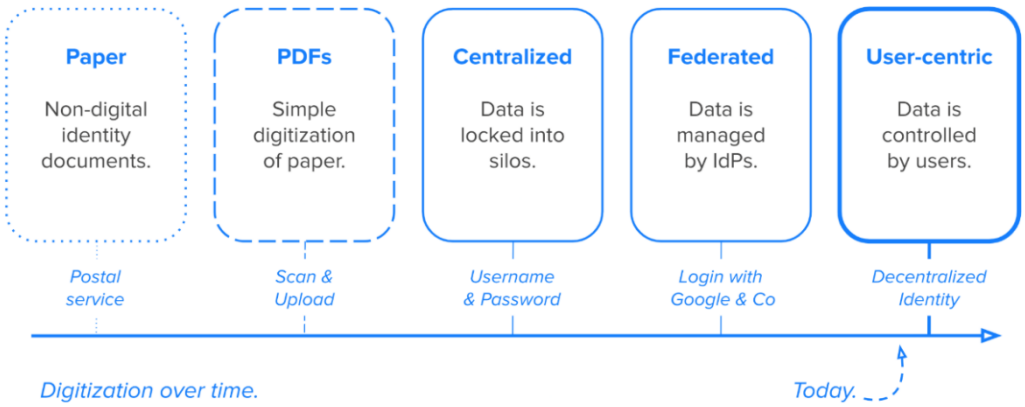


Figure 1 – Evolution of identity management (walt.id, 2022)

SSI was designed around user autonomy and privacy. SSI empowers users with full control over their personal data on who gets access to their information and under what conditions. Accomplished through the wiring together of cutting-edge technologies working together in accomplishing a secure, decentralized identity ecosystem (walt.id, 2022).

- Decentralized Identifiers (DIDs) are globally unique identifiers initiated, owned, and controlled by the user. Unlike the traditional means of identity, like email addresses or social security numbers, a DID functions independently of any centralized registry, hence offering a foundation for ensured, confidential, and authentic digital interaction.

- VCs are digital replicas of traditional credentials—passports, driving licenses, and degrees. These credentials are cryptographically secure, and verification can be done independently, ensuring that the information contained within them is correct and reliable. VCs are stored in digital wallets and can be shared at will with third parties; therefore, this offers the end-user fine-grained control over personal data.

- Digital Wallets: This is a secure application to store and manage DIDs and VCs. Digital wallets thus allow the user to communicate with the different services involved and share their credentials where necessary. A digital wallet is quite high security in that classified information cannot be accessed by unauthorized parties.

- Trust Registries: They can be considered the sources of trust which attest to the genuineness, validity, and integrity of DIDs and VCs. The trust registries are at the core of integrity within an SSI ecosystem, ensuring that only verified data is shared during digital interactions.

- Cryptographic keys: Keys are fundamental to the security of SSI and can be used for encryption, decryption, and digital signatures. They ensure that all communications and data exchanges are secure, and that the identity data cannot be tampered with.

SSI has a variety of benefits compared to traditional identity systems (walt.id, 2022):

- It provides improved privacy and security by giving users control over their data and minimizing data disclosure, thereby reducing the risk of data breaches or unauthorized access. Besides, SSI eliminates the use of passwords, which turn out to be the weak point in traditional systems.

- **Better User Experience:** It allows users to share the credentials without a hassle, avoiding typically annoying processes like form filling or password management at the very beginning. This results in a much more intuitive and streamlined user experience.

- **Fraud Prevention, More Trustworthy Interaction:** Because SSI credentials are cryptographically secure and verifiable, they drastically reduce the cases of identity theft and fraud. Therefore, an organization can be very confident about the validity of the credential presented to it, making digital interaction more trustworthy.

- **Compliance with Regulatory Frameworks:** By design, SSI intrinsically aligns with privacy regulations like the GDPR, simply because it was designed to empower the individual with control over their personal data and its dissemination.

Even though the usage of SSI grants many benefits, there are some obstacles of its implementation. The authors distinguished three (Alethio Preukschat and Drummond Reed, 2021) :

- Building out the new SSI ecosystem.
- Decentralized key management.
- Offline Access.

Building out the new SSI ecosystem: Base infrastructure for SSI is already under development, it is still in its infancy and hence not yet ready for large-scale deployment on the internet scale. Further substantial work is required to mature this infrastructure so that it can be reliably deployed at large scale.

Decentralized key management: While Self-Sovereign Identity fundamentally revolves around cryptographic keys, especially key pairs with the private keys directly in possession of the SSI identity owner in their personal digital wallet, misplacing these private keys would imply an absolute loss for that owner of digital identity. Effective key management has been one of the most complicated hurdles towards wide diffusion of cryptography and PKI. The challenge has

been so overwhelming that many experts are convinced that it is amenable to good management only by large organizations or centralized service providers, such as banks and government agencies. Development of cryptocurrencies, however, has lately paved the way for disruptive decentralized key management, and the birth of SSI is giving further impetus to research in this direction. This is hence a severe obstacle to SSI gaining more widespread acceptance in the market.

Offline Access: SSI relies on digital credentials shared over a network, but there are many scenarios where individuals need to prove their identity without access to the internet or a digital device. For example, the Canadian Mounted Police may have to validate a driver's license in far northern regions where access to the Internet might not be available. Thus, SSI solutions must be offline or in conditions of sporadic or low-quality connectivity. This is an enormous engineering challenge that SSI architects are working on at this moment, but far from resolved. Moreover, existing SSI infrastructure lacks the interdisciplinary competencies needed by a complex domain like identity.

2.6 Data space architecture

With the knowledge we gather from what is a data space, its design and to the supportive regulations, now we need to start thinking about the architecture, the building blocks and its implementation.

Various dataspace-related organizations provide specifications and standards to follow, reference architecture and minimal implementations. The latter are not intended to as the only implementation of their respective frameworks, but rather to showcase the possibilities and provide reusable software components and tools. There is no single best solution to establishing a dataspace, but adherence to shared standards is necessary to enable interoperability. Certain high-level technical components are common in all dataspaces but in general, a dataspace solution may consist of various elements (Reiberg et al., 2022):

- **Asset Provider**: An individual or organization offering an asset that it holds/operates.
- **Asset Consumer**: An individual or organization wishing to acquire/use an asset.
- **Compliance Services**: Services that validate all the other components of the system and ensure interoperability.
- **Identity Services: Services that provide and manage credible identities and enable trust among the system participants.**

- **Catalogue**: A registry of assets allowing providers to publish their assets and consumers to search for available offers.

- **Data Exchange**: Services that handle the transaction between provider and consumer, including contracting, logging, and data transfer.

In the following sections we will address in deep the reference architectures, International Data Spaces (IDS) and Gaia-X and how they differ from one another.

2.7 International Data Spaces (IDS) Reference Architecture

The International Data Spaces initiative is an industry-wide strategic project for secure and sovereign data exchange. IDS was launched by the Fraunhofer Society, supported by the German Federal Ministry of Education and Research, for the standardization of data sharing and the guarantee of data sovereignty. IDS pursues the creation of a data economy that would effectively and securely open data sharing for further innovation and cross-sector collaboration. Besides, it will create tasks involving proper reference architecture for secure data exchange, a certification mechanism regarding trustworthiness, and the establishment of an open and interoperable ecosystem for data sharing (Otto et al., 2021).

The IDS Reference Architecture Model is the spine of the IDS initiative, providing an all-inclusive framework that describes roles, interactions, and security mechanisms involved in trusted data exchange. This architecture secures the sovereignty of data, their interoperability, and compliance with legal and regulatory requirements.

- **Core Participants**: Data Owners, Data Providers, Data Consumers, and Data Users are the IDS ecosystem's core participants. Each participant is associated with specific roles that come into play and interact within the IDS framework to realize secure data exchange.

- **IDS Connector**: This is one of the major building blocks of the IDS architecture. An IDS Connector enables secure transfer between participants. This means that, in respect to a previously established usage policy and conditions of use, it guarantees the exchange of data under continuous data sovereignty.

The figure 2 illustrates an IDS Data Space.

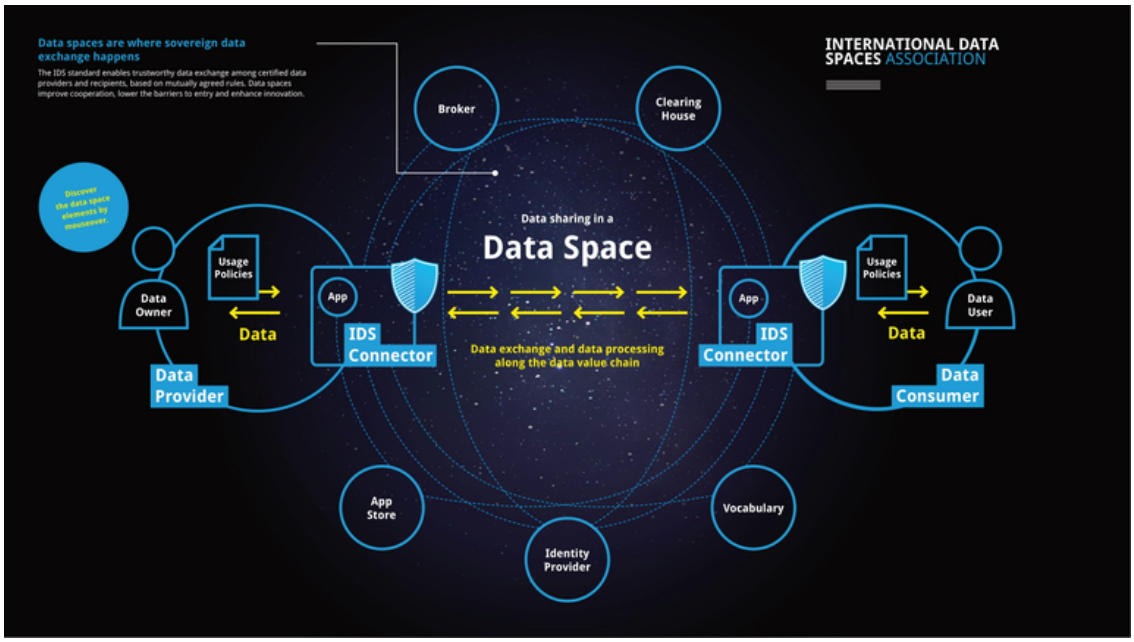


Figure 2 - Data space (International Data Space Association (IDSA), 2024)

The Certification is one of the key components within the IDS framework, ensuring that all participants and components adhere to very high standards pertaining to security and trust. The process of certification entails the following (Otto et al., 2021):

- **Organizational Certification**: This confirms whether the participating organizations have appropriate security and management processes in place.

- **Technical Certification**: Certification that the technical components, like the IDS Connector, will adhere to very strict security and functionality requirements.

In the IDS certification framework, it includes several security profiles and assurance levels from which participants can choose the appropriate level of security to bind with.

- **Interoperability** has been pinpointed as one of the success factors in the IDS initiative. In this respect, the IDS Information Model relies on standardizing the specifications of the Semantic Web and RDF to provide a harmonized approach for data description and exchange. It enables easy understanding and use of data across different systems and organizations.

- **Data Exchange**: The IDS supports the use of REST APIs and Linked Data principles to facilitate smooth data exchange. The IDS Connector enables secure, peer-to-peer data transfer to ensure that data is always available at the owner's side.

- **Shared vocabularies**: The IDS Information Model enables the provision of domain-specific vocabularies, which uniquely define a common understanding of data concepts and thus facilitate the efficient sharing and integration of data.

- **Facilitating data sovereignty**: A core primary goal of IDS is that data owner should be able to assure and enforce usage policies of their data. IDS Usage Control Language allows participants to describe their terms and conditions for data usage, hereafter enforced by the IDS Connector.

- **Policy Specification**: IDS allows definition of technology-independent, human-readable policies, in acclimating data usage obligations and restrictions.

- **Applying these policies**, it converts them into a form readable for machines and lets the IDS Connector enforce compliance with terms of use concerning data.

The IDS framework enables the execution of additional software (IDS Apps) in a secure runtime environment. These apps process data while meeting the most stringent security and isolation requirements.

- **Security Requirements**: The IDS Connector realizes a basic set of security measures, which means identity and access management, system integrity, data confidentiality, etc. Advanced trust profiles then realize further advanced security guarantees based on hardware-based trust anchors.

- **Runtime Environment**: IDS Connectors provide a standardized API and security framework inside which an IDS App can be run, therefore guaranteeing the compatibility and interoperability of a multitude of systems.

2.8 Gaia-X Reference Architecture

Gaia-X is a European initiative aimed at creating a federated, secure data infrastructure that respects data sovereignty and becomes an innovation driver in the European data economy. With the offering of an empowered system of relationships in the form of different vendors of cloud services, operating in an unquestionable and fair ecosystem, Gaia-X can look to provide a safe place on the web where people would really be in control of their information. Gaia-X empowers trusted, decentralized, digital ecosystems. Its mission is to develop the policies, rules, specifications and a verification framework that, based on European values, will become de facto the standard for digital sovereignty (Gaia-X Association, 2024a).

The key feature of Gaia-X is: (Gaia-X Association, 2024a)

- **Federated Data Infrastructure**: Gaia-X interconnects different cloud service providers; through this interconnection, sharing of data is enabled, while keeping the

uninterrupted right of users to dispose of their data concerning access to it and the possibility of its use.

- **Open-Source Implementations**: Gaia-X promotes and provides open-source implementations of its specifications to let more transparency and collaboration happen.

- **Qualification Authority:** The project serves as an independent qualification authority for the Gaia-X Label, particularly the Basic Conformity certification.

Data spaces are at the core of Gaia-X, meaning digitally mapped relationships between trusted partners with bindingly high standards in data storage and sharing. They allow for the integration and cooperation of actors along a value chain—for example, suppliers and OEMs in manufacturing or health providers in the medical sector. A data space is an economy of data that makes possible innovation and value that cannot be achieved by discrete data sets. The core aim of Gaia-X is to establish a European data economy by establishing the conditions for data outbursts with the help of developed data spaces. Different domains covering public and private sectors will organize a data space, facilitating collaboration and using high-quality data for competitiveness in a digital world (Gaia-X Association, 2024a).

The image bellow demonstrates Gaia-X framework where we also visualize 3 pillars - compliance, data exchange and federation.

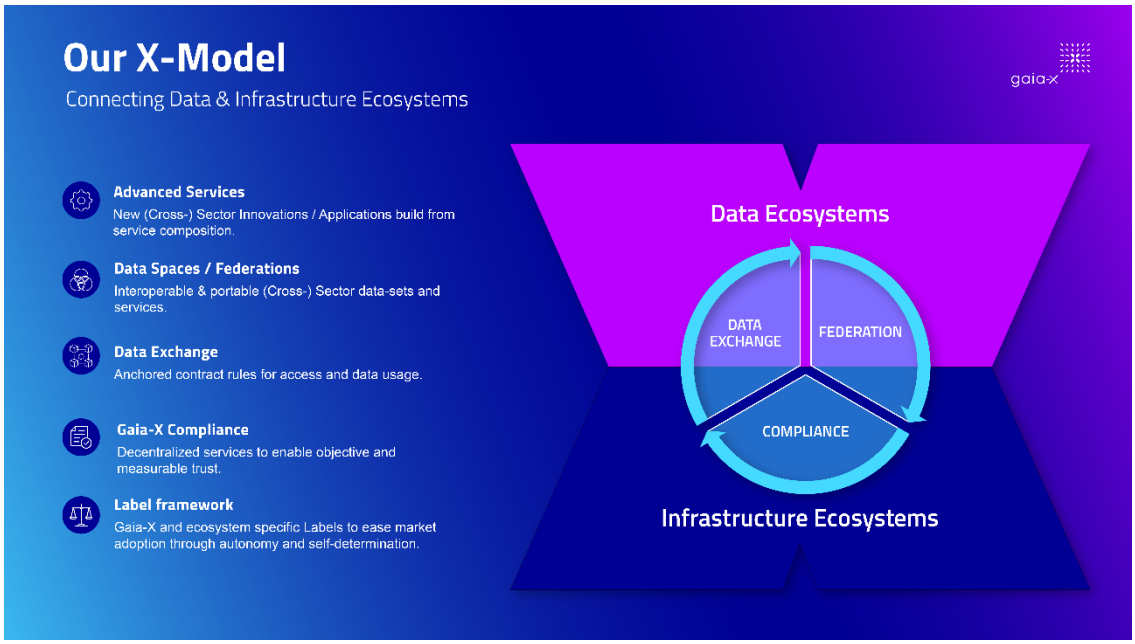


Figure 3 - Gaia-X X-Model (Gaia-X Association, 2024a)

The basic assets of the Gaia-X ecosystem are Nodes, Services, Service Instances, and Data Assets: (Otto et al., 2021)

- **Nodes**: They are computational resources that range in scale from Edge Devices via Virtual Machines to Data Centers; they form the infrastructure on which services can be deployed.

- **Services**: Cloud offerings that are offered by a Service Provider, becoming a Service Instance when deployed on a Gaia-X Node. Services can interoperate with each other and across Nodes, enabling flexible combinations.

- **Data Assets**: Datasets exposed through Gaia-X Services, which are physically hosted on Nodes and shared between participants. They may be used within the federated ecosystem to compose data spaces.

Participants, usually businesses, can be providers, consumers, or both. Every asset and participant shall create a so-called Self-Description—a structured metadata record—including the information of ownership and usage policy, among other attributes. Self-Descriptions are key in onboarding and establishing trust within the ecosystem The Gaia-X Framework is hence based on three pillars: **Compliance, Federation Services, and Data Exchange**. On one hand, compliance services guarantee decentralized trust; on the other hand, Federation services enable interoperability and portability. The Data Exchange services manage transactions, contracting, and access control. It also embeds Gaia-X Digital Clearing Houses as execution nodes for compliance services within the framework (Otto et al., 2021).

According to Van Der Schaaf et al. (2022) , Gaia-X consists of four main blocks: Identity and Trust, Federated Catalogue, Compliance and Data Sovereignty Services.

Identity and Trust Services: They are responsible for handling the participants' identities and thus create trust in the ecosystem (Van Der Schaaf et al., 2022).

Federated Catalogue: It mostly handles how the participants fine each other's (Van Der Schaaf et al., 2022).

Compliance: Compliance includes a harmonized legal framework governing the relationship between service providers and consumers in terms of policies, metering, and billing, and the rights and obligations of the parties. It is in these service interactions that self-descriptions are instrumental in ensuring adherence to these policies (Otto et al., 2021).

Sovereign Data Exchange Services allow participants to control the use of their data by asserting the principle of data sovereignty (Van Der Schaaf et al., 2022).

2.9 Gaia-X and IDS

Both initiatives while targeting the cultivation of data sovereignty, trust, and interoperability in the digital economy, the share similar goals but differ significantly by approach, architecture, and areas of focus. The following section explains the core differences between Gaia-X and IDS and points out what is unique in these contributions to the data ecosystem (Otto et al., 2021).

From **architectural perspective** Gaia-X is focused mainly on the infrastructure for federated and interoperable cloud structures that guarantee data sovereignty and support smart data applications in many industrial sectors. Gaia-X includes the infrastructure ecosystem and the data ecosystem, which interconnect through federation services. This shall eventually allow for seamless exchange of data and services within a federated cloud environment. On the other side, IDS is focused on the creation of a secure data space, which would enable organizations to manage and exchange data independently but remain in control of their data. The IDS Reference Architecture Model (IDS-RAM) gives an idea of how safe and trustworthy data exchange can be conducted within the confinements of the business ecosystem. IDS emphasizes data governance, usage control, and standardization of data sharing protocols (Otto et al., 2021).

In terms of core components and roles Otto et al. (2021) explains that in Gaia-X the core components the Nodes, Services, and Service Instances, which are part of either the infrastructure ecosystem or the data ecosystem. Nodes in Gaia-X represent any kind of computational resource, while Services and Service Instances describe cloud offerings that can be deployed on these nodes. Contrarily, IDS defines the participants to be Data Owners, Data Providers, Data Consumers, and Data Users. The roles identified herein are an integral part of the IDS framework; it ensures that data exchange happens securely and based on predefined policies. This makes the IDS Connectors secure gateways for data exchange, thus extending their functionality across both the data and infrastructure layers.

Still with Otto et al. (2021) while the goal of Gaia-X and IDS **is to enhance interoperability** within the boundaries of their respective ecosystems, they approach it very differently. Concretely, Gaia-X builds upon existing standards and technologies to construct a federated cloud infrastructure in support of interoperability not only at the level of data but also at the level of services. It encompasses REST API description developments and encourages the usage of JSON-LD for data serialization. In contrast, IDS focuses on common knowledge about patterns of data exchange, data formats, and sequences of interaction. In this respect, the IDS Information Model is based on specifications for the Semantic Web and RDF. It provides a sound basis for the interoperability of various systems of different organizations. IDS supports REST interactions but extends them by security and accountability attributes.

Certification and trust are an integral part of both Gaia-X and IDS. The former provides Identity and Trust services, which ensure the authenticity and trustworthiness of participants. Such services draw from concepts in the IDS—Identity Provider and Dynamic Attribute Provisioning Service—to provide a strong foundation for identity and attribute management in the Gaia-X

ecosystem. IDS has a profound process of certification for both operational environments and technical components. This can include a matrix for security profiles and assurance levels that meet Gaia-X requirements. The IDS certification framework allows for the trustworthiness of participants and the components to be deployed, hence building trust within this ecosystem (Otto et al., 2021).

Data sovereignty is a common core objective of both Gaia-X and IDS according to Otto et al. (2021). The former realizes mechanisms for sovereign data exchange by putting its owners in a position to manage and monitor usages of their data in accordance with usage policies, establishing decentralized but auditable means of logging business transactions. This is somewhat like what IDS does, as it is also interested in the usage control and enforcement of policies on how the data is being used. In this regard, the IDS Usage Control Language defines and shares terms and conditions for data usage by participants, therefore ascertaining that data shall be used according to the set and agreed-upon policies of the provider. It forms a part of the IDS framework, providing automated means for setting enforceable configurations (Otto et al., 2021).

Otto et al. (2021) explains that by assuring **data sovereignty, interoperability**, and trust in the digital economy-motivates the efforts of both the Gaia-X ecosystem and the International Data Spaces' initiatives, though they differ in architecture and approach. The key objective of Gaia-X is to establish a truly federated infrastructure that seamlessly integrates nodes, services, and instances of services, thus making flexible data exchange across different cloud environments possible. On the other hand, IDS pursues data spaces where participants maintain their independent control over the data with the guarantee of observing predefined policies. All the frameworks rely on a sound trust, compliance, and control mechanism in data use to ensure security and transparency in data exchange. Both frameworks put together represent key inputs in developing a trusted and secure digital ecosystem where innovation will have opportunities to thrive across industries.

Before beginning with the detailed review of how data connectors technically work, let's set the scene by describing the bigger picture they operate within data ecosystems. Dynamic, collaborative environments allow the flow of data between participants with innovation and operational efficiency in mind. Data ecosystems form a base architecture for both the Gaia-X and IDS frameworks as a way of structuring how secure, decentralized data sharing will be handled across industries. A data ecosystem covers design principles, governance, and how secure and interoperable interactions are ensured in both Gaia-X and IDS. The next section will examine it, helping to shed further light on the role of the data connector.

2.10 Data Ecosystems

The data ecosystem is a network of actors that provide, consume, and act as an intermediary in the exchange of data between various sources. Data spaces mediate across stakeholders in such ecosystems through building trust, interoperability, and sovereignty over data shared. The

definition of a data ecosystem goes beyond one data space to several organizations and sectors combined in the pursuit of common objectives. This collaborative approach empowers better data quality, innovation, and new business models. Data ecosystems, on the other hand, offer organizations ways to fully unlock the value of data as a critical asset—bringing together, in a secure and controlled way, data exchange from multiple sources (Boris and Hompel, 2022).

This rapid transformation of the digital landscape has thrown into consideration several concepts-data ecosystems said Boris and Hompel (2022), for example-interconnected environments in which participants such as organizations, industries, and service providers come together to share, manage, and use data. Data ecosystems are necessary to realize the full value of data by enabling cross-sector collaboration with data sovereignty and innovation. In a well-structured data ecosystem, the stakeholders enjoy frictionless data exchange, better interoperability, and increased trust-all factors that spur new business models and services. At the heart of this concept is that of data spaces, which are, by definition, mechanisms for safe and controlled data sharing within such ecosystems. Before moving on to the definition of these different types of data spaces, it is very relevant to refer to the key role that data ecosystems are playing in creating a collaborative digital environment able to support complex and highly data-driven processes in many sectors.

2.10.1 Agricultural Data Space (ADS)

ADS promotes digitalization in agriculture, considering that there is an integration of data and services coming from different platforms, improved interoperability, and data sovereignty for farmers. There are a lot of isolated solutions, not interoperable, in the agricultural domain, which add to redundancy and general inefficiency. ADS resolves these by providing a harmonized framework for data sharing and integration. This data space enables the effective support of all agricultural business and work processes, allows for new business models, and provides more transparency from seed to plate. The ADS concept is derived from the Fraunhofer lighthouse project "Cognitive Agriculture" and is supposed to help overcome current obstacles to a broader diffusion of digital technologies in agriculture (Boris and Hompel, 2022).

The figure 4 illustrate a agricultural data space.

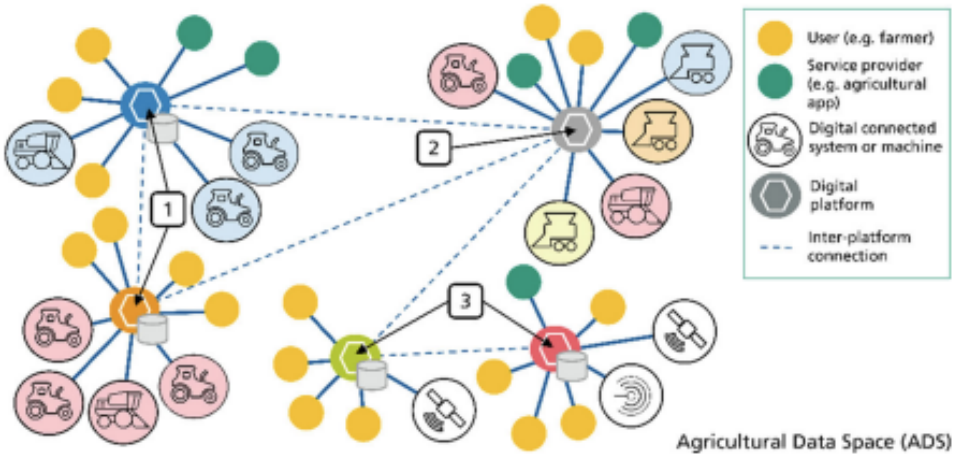


Figure 4 - Agricultural Data space (Boris and Hompel, 2022)

There are three application scenarios that would most of all benefit from an ADS:

2.10.1.1 Sustainable Management of Nutrient Cycle

Another important application of ADS would be in the realm of sustainable nutrient cycle management. This information can be obtained from various data sources on tracing and managing the flow of nutrients on the farm, thus helping farmers to make optimum use of any type of fertilizers with minimum damage to the environment. For example, with knowledge of soil quality data, crop yield data, and weather forecast information, farmers can decide on the optimal time and quantity of fertilizer application. This will not only increase crop productivity but also reduce nutrient runoff into water bodies. A farmer can utilize the ADS to participate in data-intensive applications for yield comparison, equipment efficiency evaluation, and food chain transparency (Boris and Hompel, 2022).

The figure 5 demonstrates how sustainable management of nutrient cycle works.

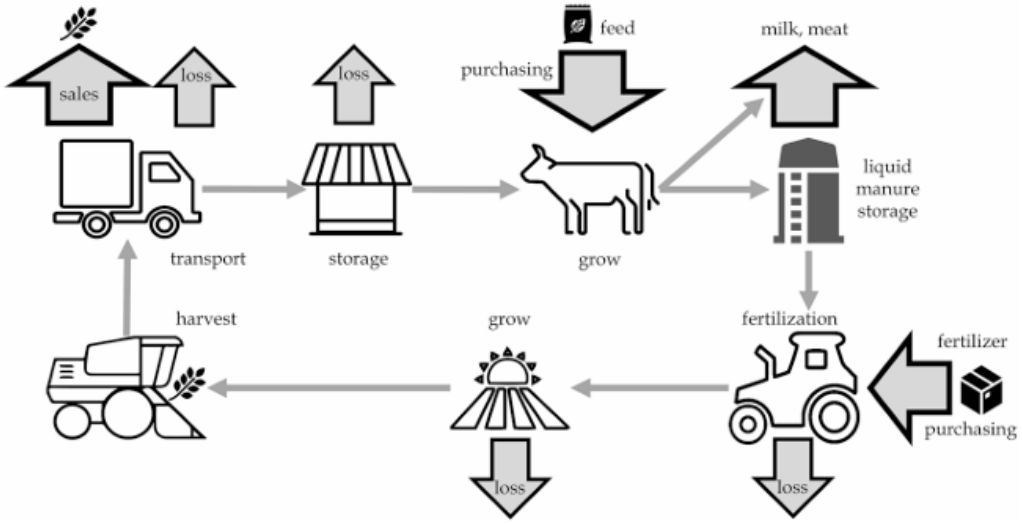


Figure 5 - Sustainable Management of Nutrient Cycle (Boris and Hompel, 2022)

2.10.1.2 Governmental Platforms

Another key use case would thus be government platforms managing data pertaining to agricultural regulations and subsidies. Public authorities would be able to develop a platform for processing data provided by farmers for acquiring regulatory information and applying for subsidies. For example, a government platform would be able to collect data on farming practices with respect to compliance with environmental regulations and, on that basis, allow for relevant subsidies to be granted. By making this space available to these platforms through a link from non-governmental IT systems and digital platforms, ADS facilitates interoperability and the use of data that is already available to farmers and companies working in agriculture. This will assist farmers in complying with their legal obligations while remaining in control of their data (Boris and Hompel, 2022).

2.10.2 Medical Data Space

The purpose of MDS is to enhance sensitive medical data exchange among health providers, researchers, and patients, subject to very high trust and security conditions. The heterogeneous context of a healthcare sector, where the degree of digital maturity varies from one provider to another, also calls for. MDS must seal the gap by enabling smooth data exchange among the concerned organizations and at the same time supporting new healthcare concepts—data-based, such as management or healthcare, and disease-specific, integrated care or precision medicine. An MDS architecture is made up of collaborative, interactive information systems or "apps" that encapsulate functions and data to enable user interaction and to manage healthcare processes (Boris and Hompel, 2022).

There are several scenarios where the medical data space provides value.

2.10.2.1 **Health and Disease Management**

One of the major fields for using MDS is health management and disease management regarding chronic diseases, such as diabetes. Concerning this, the SALUS project is funded by German *Innovationsfonds*. In the case of glaucoma management, a novel approach has been realized within the scope of SALUS. Patients monitored intraocular pressure at home, using an electronic health record app, which reduced the burden of frequent hospital visits. The data is collected via an app from a consumer-operated measurement device and reviewed by an ophthalmologist through a web-based information system. The measuring process can be conducted continuously with timely interventions, improving patient outcomes and quality of life. In integrating patient-generated data into clinical data, MDS supports comprehensive strategies for disease management (Boris and Hompel, 2022).

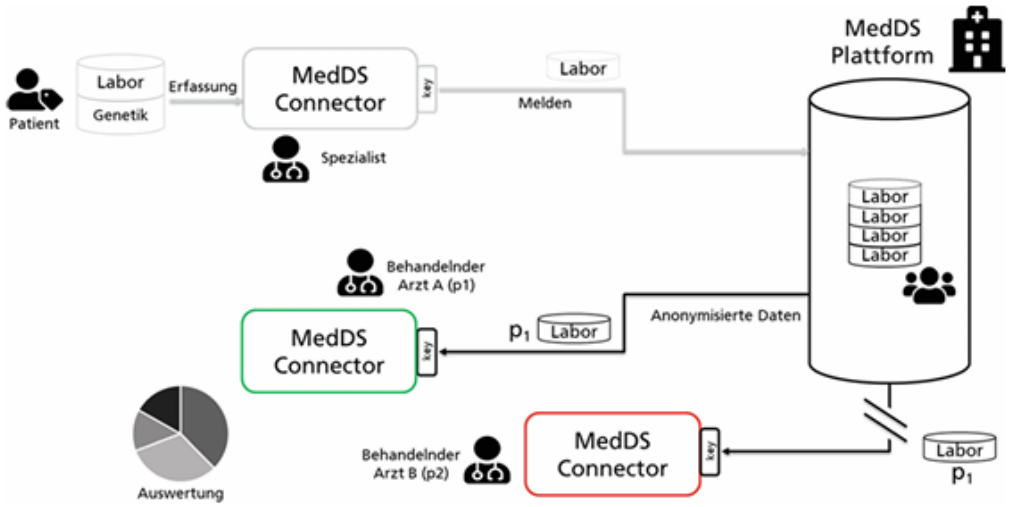


Figure 6 - **Health and Disease Management** (Boris and Hompel, 2022)

2.10.2.2 **Precision Medicine**

Another critical application of MDS is in precision medicine, which requires medical treatment to be tailored according to the characteristics of each individual patient. In the case of cancer treatment, for instance, one can use genetic data of a patient's tumor to identify which mutations exist in the tissue and select the most efficient targeted therapies. In health care, developers of personalized therapy plans have improved intervention efficacy by reducing side effects by integrating genetic, clinical, and lifestyle data through MDS. The integration of genomic, clinical, and lifestyle information further allowed better care for the patient and contributed to medical research through the delivery of invaluable data needed in developing new therapies or, at times, even understanding the mechanisms underlying some diseases (Boris and Hompel, 2022).

2.10.3 Energy Data Space

An Energy Data Space allows for secure, sovereign data exchange along the entire energy value chain, hence supporting the sector in achieving decentralized and renewable energy systems. In addition, it offers secure data exchange and thus boosts innovative development of smart energy services and solutions based on GAIA-X infrastructure. This is an important initiative to help integrate renewable sources of energy, optimize energy efficiency, and develop more reliable and sustainable energy systems. It foresees a structured environment for data sharing, thus enabling the real-time supply and demand balancing as required in efficient smart grid operations (Boris and Hompel, 2022).

According to Boris and Hompel, (2022), the data can be seamlessly integrated from different sources of energy, grid operators, and consumers, thus underpinning predictive maintenance strategies and proactive energy infrastructure management. In this way, data from sensors of the wind turbines will help in planning service and reducing downtime while increasing operational efficiency. In addition, it underpins new business models like peer-to-peer energy trading or energy as a service by providing a safe platform that makes every consumer able to purchase and sell energy directly, enables collaboration between actors to increase innovation and efficiency of the energy sector.

Still with (Boris and Hompel, 2022), the Energy Data Space shows huge potential for fostering innovation in the energy sector. In this respect, it will be of major significance for coping with the increasing digitalization and decentralization of the energy system and ensuring that it is securely and efficiently operable. Further developments could also include more advanced data analytics and AI-driven insights that would optimize energy production, distribution, and consumption. The Energy Data Space will help achieve sustainability targets and enhance overall efficiency and resilience with respect to the energy system, ensuring a sustainable and competitive energy future, by establishing a collaborative setting for data exchange (Boris and Hompel, 2022).

There are several scenarios where the energy data space provides value.

2.10.3.1 **Smart grid management**

Smart grid management is one of the critical use cases for an Energy Data Space. Integrating data from the various sources of energy—from renewable energy producers through grid operators to consumers—balance supply and demand in real-time with the aid of the Energy Data Space. For instance, it can foresee energy consumption patterns by using data recorded on smart meters, which grid operators then use to adapt energy distribution accordingly. It can, therefore, help to have efficient grid operation, reduction of energy losses, and reliability in the power supply. In that respect, this integration of real-time data avails in Energy Data Space facilitates the integration of renewable energy sources for a sustainable and resilient energy system (Boris and Hompel, 2022).

2.10.3.2 Predictive Maintenance

Boris and Hompel (2022) describes the predictive maintenance in the energy sector as a method of maintaining equipment before it fails, based on projections obtained from analyses of various data. For example, from the sensor data on its turbines, a wind farm could detect anomalies and thereby plan maintenance in advance of a breakdown. This way, predictive maintenance increases not only the effectiveness and reliability of energy production but also reduces maintenance costs and downtime. Levelized information is shared securely across the different stakeholders, from maintenance teams to equipment manufacturers, in the Energy Data Space. That enables a collaborative, very effective maintenance strategy. This use case illustrates how data can improve operational efficiency and prolong the lifetime of key energy infrastructure.

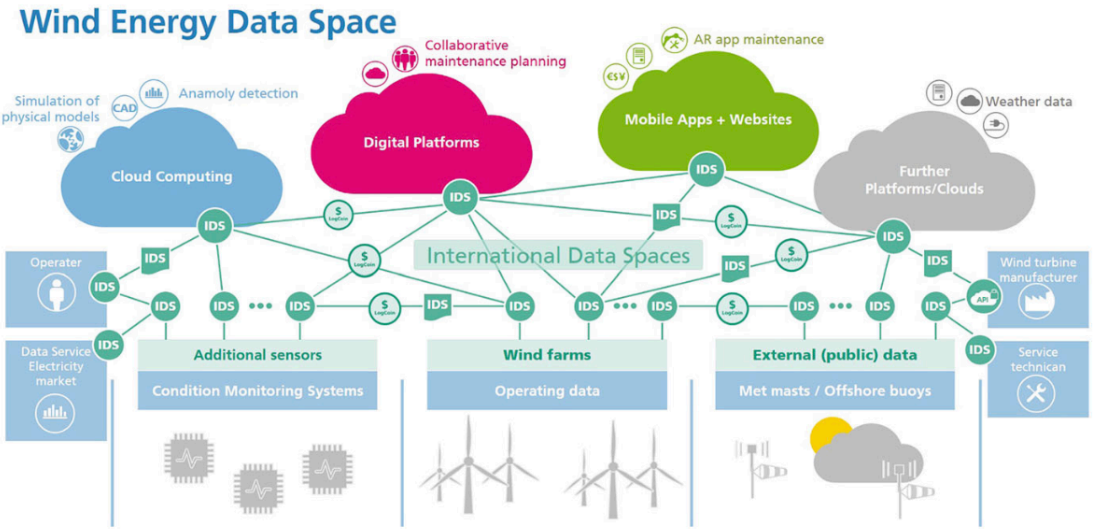


Figure 7 - Wind Energy Data Space (Boris and Hompel, 2022)

Predictive maintenance applications deliver valuable data to wind farm operators and maintenance service providers, but also to wind turbine manufacturers and component suppliers for product development to improve the efficiency and fault resistance of turbines. This data is also useful for consulting services and insurance companies in finding out operational dynamics and potential risks. This data becomes enriched with expert knowledge related to fault identification and damage prediction in the process of analyzing the anomaly detection data by experts. The integration of expert analysis also establishes new business relationships and models related to operation and maintenance processes. The possibility of collaboration and extensive benefits arising from a well-designed Energy Data Space is evidenced by the spectrum of different functionalities performed by various participants within the data ecosystem (Boris and Hompel, 2022).

2.10.4 Industrial Data Space

Industrial Data Spaces are tailored to guarantee secure and sovereign data exchange for industrial sectors in manufacturing, logistics, and supply chain management. In the same way, data spaces solve the problem of integrating data originating from several isolated systems so that companies truly benefit from digitalization while staying in control over their data. IDS is a concept for an architecture for data sovereignty, security, and interoperability that allows the sharing of data providers' information without running the risk of unauthorized access or misuse. This secure environment for business model innovation supports increased operational efficiencies and collaborative innovation throughout the value chain of industry (Boris and Hompel, 2022).

2.10.4.1 Collaborative Condition Monitoring (CCM)

Collaborative Condition Monitoring is one of the huge use cases in IDS. Sharing sensor data and machine data among a couple of organizations is involved in this use case for condition monitoring equipment and maintenance needs prediction. For instance, in a manufacturing network, real-time data from production machinery can be aggregated to analyze early signs of equipment failure. By making this information available to manufacturers of machines and providers of maintenance services, predictive maintenance strategies that minimize equipment downtime and prolong its life can be devised. Only through collaboration on this level will operational efficiency improve, but it will also improve the reliability and safety of the industrial process (Boris and Hompel, 2022).

2.10.4.2 Smart Factory Web (SFW)

Another critical application of IDS in Industry 4.0 is Smart Factory Web, where factories may be in different places and are connected for the purpose of production data exchange and optimization of manufacturing processes. For example, factories within a global supply chain could coordinate their production schedules with respect to inventory levels for an appropriate supply and demand balance. Thanks to the safe data exchange functionalities of IDS, these factories can reduce scrap, bring down their costs, and react dynamically to market changes. The SFW shows how IDS can transform traditional manufacturing into interconnected, adaptive production networks and take a quantum leap in productivity and competitiveness. interconnected, adaptive production networks and take a quantum leap in productivity and competitiveness (Boris and Hompel, 2022).

2.10.4.3 Digital Twins

Starting with Industry 4.0, DT has been one of the most critical technologies, enabling the digitalization of physical assets in terms of their digital counterpart. Advanced monitoring and predictive capabilities of the digital twin became a reality through real-time data synchronization. In this respect, it enables the transformation of the industries' way of working. DTs can attain enhanced potential toward interoperability, data sovereignty, and secure data sharing by integration with the concepts of data spaces and Gaia-X (Zenza, 2024).

Zenza (2024) describes the digital twins as the virtual replicas of physical entities. This could pertain to products, systems, or processes. It mainly has three parts in the twin design: the physical asset, its digital counterpart, and the data connections which enable real-time synchronization between the two. It provides continuous updating and insights through this dynamic interplay, thereby enabling predictive maintenance, optimized operations, and informed decision-making.

(Zenza, 2024) several unique characteristics of digital twins:

Real-Time Synchronization - Between the digital and physical entities, continuous data exchange keeps the virtual models updated and high-fidelity models of physical assets would hold a truthful and elaborate description of geometry, behavior, and operational state.

Bidirectional interaction - These enable feedback loops from virtual-to-physical, whereby simulations can affect real-world operations. Among the several advantages that Digital Twins offer is the following:

- **Monitoring**: The performance and health conditions of assets are always made visible.
- **Predictive maintenance**: Early detection of impending failures and schedule optimization for maintenance.
- **Operational Efficiency**: The allocation of better resources and processes, with reduced downtime.
- **Innovation and Prototyping**: Faster development cycles were enabled because of the way virtual prototyping and testing could be done faster.

2.10.5 Mobility Data Space

Mobility Data Space shall transform the transport sector by creating an open and interconnectable data space with access to real-time traffic and mobility data. The initiative combines data from local, regional, and national platforms into one point of access for all mobility information. With its decentralized architecture, data providers will be able, through the Mobility Data Space, to manage their data and establish any usage conditions that ensure sovereignty and engender trust in the data. It provides for the establishment of a framework to

support the development of intelligent transport systems that can optimize traffic flows, enhance safety, and reduce environmental impact (Boris and Hompel, 2022).

2.10.5.1 Real-Time Traffic Management

One major use case for the Mobility Data Space would be real-time traffic management. In a Mobility Data Space, data can be integrated from several sources, like public transport providers, navigation services, and fleet operators, providing detailed and correct information on traffic. Cities could adjust traffic light timings dynamically against the backdrop of the current traffic situation, thus optimizing overall traffic flow and avoiding congestion in cities. It could also, therefore, feed real-time data to drivers and passengers to take the best routes, bypassing congestion, which would increase urban transport efficiency and sustainability (Boris and Hompel, 2022).

2.10.5.2 Multimodal Commuter Information Services

Another key field of application will be in delivering multimodal commuter information services. The Mobility Data Space can bundle data coming from several modes of transport, such as buses, trains, or bike-sharing systems, for perfect and efficient commute planning services for passengers. For example, commuters will receive real-time updates and suggestions for the quickest routes or the most cost-effective routes using different modes of transport. Such an integrated approach offers maximum convenience and efficiency for commuters, therefore serving to promote the use of public and sustainable transport modes for mitigating urban congestion and environmental impacts. In other words, it generally means that the ecosystems of both Gaia-X and International Data Spaces work for data sovereignty, interoperability, and trust in the digital economy, though they do so with different architectures and approaches. The main emphasis in Gaia-X is on the federated cloud infrastructures which support both the infrastructure and data ecosystems through nodes, services, and service instances (Boris and Hompel, 2022).

Having considered all the important components of both Gaia-X and IDS, let's embark upon describing the data connector, one of the crucial enablers in secure data exchange and interoperability within these ecosystems. The data connector provides the key interface that securely bridges participants of a data space in a way that allows for data flows according to policies and standards. Data Connector: The data connector enables data sovereignty and will be discussed in the next section as the key to secure and frictionless data exchange in a decentralized environment for both providers and consumers.

2.11 Data Connector

According to the report provided by Giussani Giulia and Steinbuss Sebastian (2023), a Data Connector is a tool used to share data in a secure way with effective communication and exchange in data spaces ensuring that the participants can freely share data. They act as nodes in the data space providing sovereignty. It realizes two important aspects, providing an Application Programming Interface (API) to other participants in a data space to achieve interoperability and implementation of the policy enforcement mechanism and a common baseline for cybersecurity.

In the image below we can see where the data connector is situated in a data space.

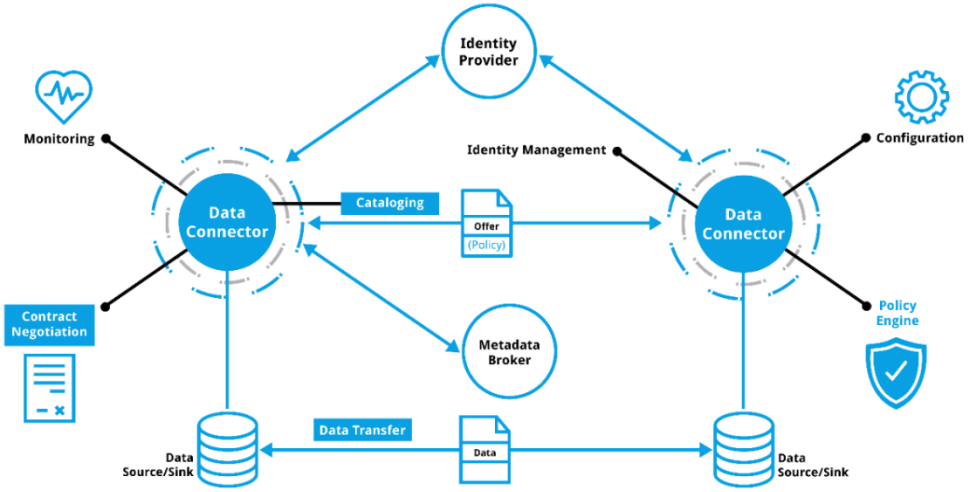


Figure 8 - Data Connector (Giussani Giulia and Steinbuss Sebastian, 2023)

Giussani Giulia and Steinbuss Sebastian (2023), mentions in the report that the technical interoperability is a major requirement in data spaces. Data connectors should be realized based on specifications and standards rather than relying on singular implementations. The general interaction between the connectors for the description of data assets and the related endpoints must be addressed including the definition of policies for access control and usage control, followed by the negotiation of those policies and contracts.

In the table below we can see a list of the available data connector implementations. The one mark with a “x” means that they are open source and “planned “means that they are still in the process to become open source (Giussani Giulia and Steinbuss Sebastian, 2023).

Table 1 – Available data connectors (Giussani Giulia and Steinbuss Sebastian, 2023)

Name of the connector	Open source
AI.SOV connector	
Dataspace Connector (DSC)	x
ECI IDS Connector powered by TNO	
Eclipse Dataspace Connector (EDC)	x
EGI Datahub connector	x
IDS Integration Toolbox	x
GAIAbiX. IDS. BasicConnector	
IIOC (Intel IONOS Orbiter Connector)	Planned
Kharon IDS Connector	
MPAD-C	x
OneNet Connector	Planned
Silicon Economy EDC	x
sovity Connector-as-a-service	
Tech2B SCSN Connector	
Telekom DIH connector TeraLab Connector	x
TNO Security Gateway (TSG)	x
Tritom Connector	
True connector	
Trusted Connector	
Trusted Supplier Connector (TSC)	
VTT DSIL Connector	
WeTech Smart Data Connector	

2.12 Internet Of Thing (IoT)

2.12.1 Introduction

There are different definitions of IoT, but it can be defined as a system that interconnects computing devices, be they computers, machines, people, animals or objects, and within which they can communicate and transfer data without any human intervention. The IoT allows targets to be monitored and acted upon remotely from the network, enabling full integration between computer systems and the physical world. IoT is used to power topics such as smart grids, smart homes, traffic control, parking and transportation networks, among many others. There are low-cost applications and devices for everything we can think of, boosting all kinds of communication, surveillance, transaction and remote action. (Coelho, 2017)

2.12.2 Architecture

The IoT architecture consists of several layers (Coelho, 2017):

- **Hardware** - The connected devices must integrate the application infrastructure, namely by being compatible with the architecture.

- **Data model** - Data collected from devices and other sources can be structured or unstructured. This means that ideally it will be data with a known and well-defined structure, but that an IoT architecture can deal with unstructured data resulting from sources where the data models will not be perfectly defined.

- **SDK** - Each device will run code generated by the central application, designed to be entered into a remote system.

- **Communication** - As we have seen, this type of application presupposes some kind of connectivity between the devices and the central system.

- **Backend Application** - One of the main parts of an IoT system is the central application, which not only has the task of dealing with the many devices that are connected to it.

- **Data Processing** - Data processing systems need to be able to process and correlate large volumes of data, as well as integrate with existing data processing and exploration technologies on the market, enabling them to be processed, explored and visualized.

- **Business Applications** - At the end of the line are the business applications, which will use the services of the middleware and data processing platform to interact in a contextualized way with remote devices.

2.13 Challenges for European Data Spaces

Several challenges must be surmounted on technical, organizational, and economic levels to create data spaces in Europe (Ulrich Ahle, 2021).

Ulrich Ahle (2021) in case of technical challenges, he addresses that the data with clearly defining policies of ownership and implementing them consistently on different platforms is a difficult task. He also addresses that by making a decentralized architecture that meets the requirements of various stakeholders can be difficult.

Still with Ulrich Ahle (2021), as an **organizational challenge, he** explains that the common practices about security, privacy, and assurance should be developed against different environments of data to instill trust, but it is not easy due to different standards of governance.

As an economic challenge, Ulrich Ahle (2021) mentions establishing infrastructure for diverse data monetization approaches is crucial; however, it necessitates adaptability to accommodate various business requirements and situations and establishing strong incentives for sharing data and dispelling reservations, usually based on competitive anxieties or a lack of awareness, is indispensable for the comprehensive diffusion of data spaces.

In this chapter we talk about the foundations of data space, data connectors and the technical and practical aspects of Gaia-X framework as well of the challenges and opportunities in implementing Gaia-X in different scenarios.

In the following chapter we will examining in-depth the Gaia-X ecosystem by exploring its architecture, the operational model and the core components.

3 GAIA-X Ecosystem

3.1 Introduction

Gaia-X Ecosystem, an innovative architectural concept since the efforts to foster digital sovereignty for Europe began. Gaia-X is much more than just a technical entity; it is a strategic step toward a federated, compliant, and better-connected data space because a multitude of cloud service providers and data users can be integrated within it and provide a sound basis for secure and sovereign data exchanges. This chapter will go into depth on how the framework of Gaia-X supports the federation in data infrastructure to guarantee data sovereignty and security across borders.

First, we need to understand the architecture of Gaia-X before creating a dataspace that would follow the latter's set of directives. With deep knowledge of the framework structure and guidelines for operational work, the developed dataspace will not only respect strict European regulations but also optimally leverage the federated services to reinforce the interoperability and security of the data. The knowledge provided in this chapter will arm the stakeholders with all the necessary insights to be able to navigate effectively in the complexities of implementing the Gaia-X compliant data spaces.

By the end of this chapter, the readers will have more profound insights into how the operational dynamics are structured within the Gaia-X ecosystem and understand why its architecture stands for the core for industries working with data in a secure, compliant, and sovereign way.

3.2 The Gaia-X Ecosystem

Gaia-X acts as a driver of dynamic and interconnected digital ecosystems through the establishment of a framework for the secure exchange and integration of data and services. Based on open standards and federated principles, Gaia-X empowers different sectors and industries to cooperate effectively, thus breaking traditional walls within data silos and stimulating innovation across borders. Its architecture is envisioned to support a wide array of use cases, from smart manufacturing and logistics to health and energy management. Gaia-X delivers the trusted environment for sharing data and integrating services that will let businesses and organizations unlock new opportunities of growth and efficiency, develop innovative solutions and business models, and leverage the collective intelligence and resources available within the ecosystem itself. Value creation within Gaia-X is one of its key enablers of ecosystems under the concept of data sovereignty and control for participants. By allowing the definition of terms of sharing and use by the owners of the data, Gaia-X empowers

an organization's capability to be in control over their data assets while participating in collaborative ecosystems. This approach would not only provide higher levels of trust and transparency but also participation from the widest possible stakeholder base, especially SMEs who would otherwise exercise caution in sharing their data. Gaia-X offers an open and inclusive environment that empowers democratization in access to digital resources and capabilities, unleashing collaboration among participants who can innovate on even terms (Gaia-X Association, 2024).

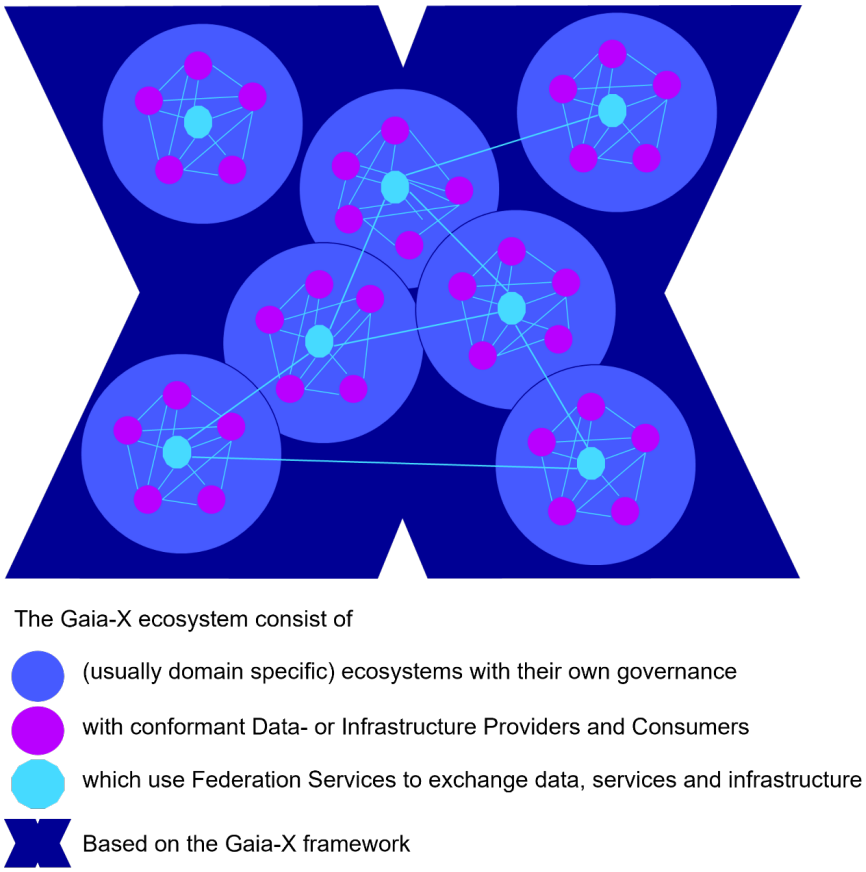


Figure 9 - Gaia-X ecosystem (Gaia-X Association, 2024).

3.2.1 Goals of Federation Services

Federation Services ensure that Gaia-X participants can have trusted and secure interactions that are seamlessly interoperable. Core services in this category include identity management, compliance verification, and data exchange, which put into action participants' interaction. This means that through Federation Services, participants are given the basic ability to work within a trustworthy and harmonized environment, whereby data and services could be reached and shared with others efficiently across multiple platforms and infrastructures (Gaia-X Association,

2024). Gaia-X Association (2024) describes the Gaia-X ecosystem as a revolutionary way in which industry-wide secure, sovereign, and interoperable data infrastructures can be developed. Embracing federation in nature, Gaia-X is a collaborative effort among cloud providers, organizations, and stakeholders with strict assurance of data sovereignty and adherence to European regulations.

Still with Gaia-X Association (2024), the architecture provides the core for dynamic data space development, leading not only to tearing down traditional silos but also innovating trust in the data-sharing ecosystems. In environments that interact with one another, businesses unlock new opportunities for growth, efficiency, and innovation. This brings us to the three critical aspects behind Gaia-X: **compliance, data exchange, and federation services**. Indeed, putting together how all these go together in support of the making of trusted data ecosystems is quite understandable. The next chapter will discuss compliance under Gaia-X, ensured through strict verification processes, how the framework allows for seamless and secure data exchange, and how federation services enable harmonized interactions across diverse platforms and infrastructures. It underlines the core features of how Gaia-X provides a solid environment for secure, scalable, and sovereign data management.

3.3 Gaia-X compliance, data exchange and federations

According to the document about the Gaia-X architecture, it has the components to address compliance, federation and interoperable data-exchange - The specifications and the supporting open-source code are defined in the Gaia-X Framework.

The figure 10 illustrates the Gaia-X ecosystem

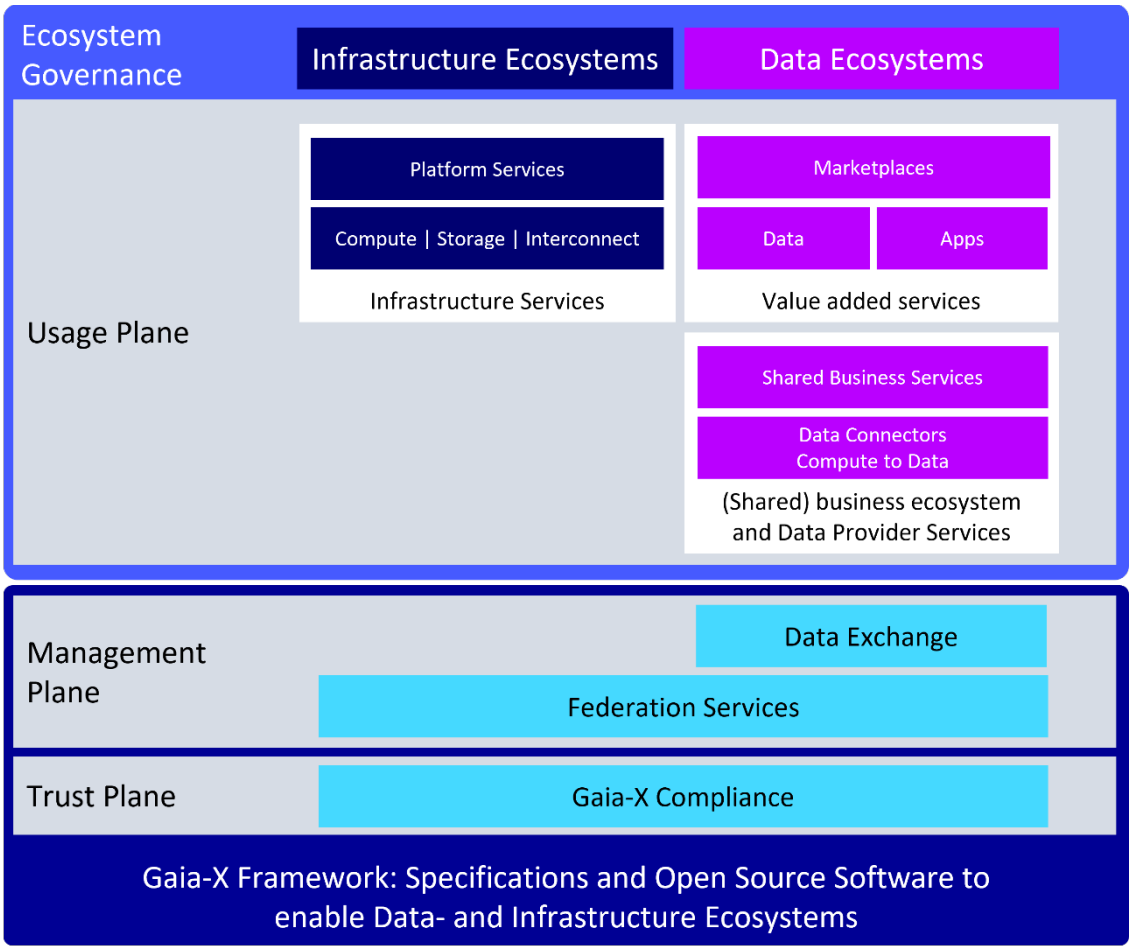


Figure 10 - Gaia-X Framework: Specifications and Open-Source Software to enable Data- and Infrastructure Ecosystems (Gaia-X Association, 2024).

In the following section we will explore in-depth all the components so that we have a better picture about the inner working of the framework.

3.4 Interoperability between ecosystems

Gaia-X defines a Trust Framework that is implemented through the delivery of two main services: the Gaia-X Registry and the Gaia-X Compliance Service. The Gaia-X Registry provides assurance about a verification and management mechanism of participant identities within the ecosystem. In contrast to this, the Gaia-X Compliance Service enforces adherence to the rules and standards contained, making sure that every one of the participants is aligned to the defined rules and guidelines pertaining to data protection, privacy, and interoperability (Gaia-X Association, 2024).

The figure 11 demonstrates the Gaia-X planes.

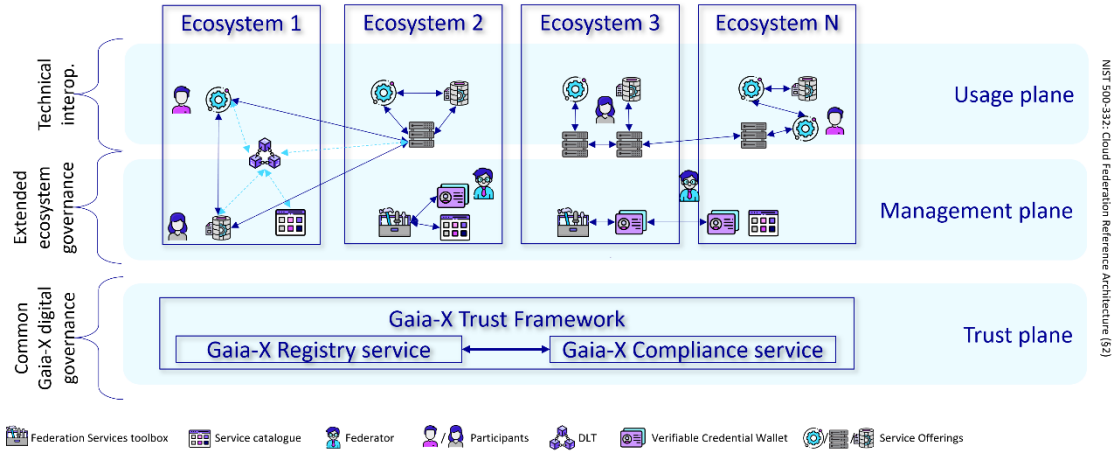


Figure 11 - Gaia-X Planes (Gaia-X Association, 2024).

Gaia-X framework is divided in 3 planes:

- Trust
- Management
- Usage

The Trust Plane is the basic layer in Gaia-X that enables secure and trustworthy interaction between participants. It realizes the Gaia-X Trust Framework, including identity verification, authentication, and authorization services. This plane creates a secure environment by checking the identity of the participants to ensure protection from unauthorized access. To add another dimension of compliance and transparency, it monitors adherence to the standards prescribed by Gaia-X and provides verifiable information about the state of compliance. This will ensure the security and reliability of the ecosystem and build trust among participants (Gaia-X Association, 2024).

The Management Plane details how participants will engage with and manage their services provided in the Gaia-X ecosystem. It features the rules of engagement and policies that participants will use to govern their activities and provides tools for service orchestration and lifecycle management. This plane allows for creating and managing federations that enable structured collaboration on certain projects. The Management Plane coordinates these activities to ensure smooth functioning and scalability, thus helping participants effectively manage resources and foster innovation (Gaia-X Association, 2024).

The Usage Plane is oriented towards the practical application of services and data from within the Gaia-X ecosystem. Through this, members will receive interfaces and APIs that allow free data exchange and integration of services, opening the way to develop solutions that create

value. This level will support the execution of use cases through service and data set combination to build new solutions and business models. The Usage Plane also provides the participant with access to service performance monitoring tools for deriving maximum value from the interaction and driving digital transformation (Gaia-X Association, 2024).

3.5 Gaia-X Conceptual Model

The Gaia-X Conceptual Model is an architectural blueprint of a federated, secure, and interoperable data ecosystem.

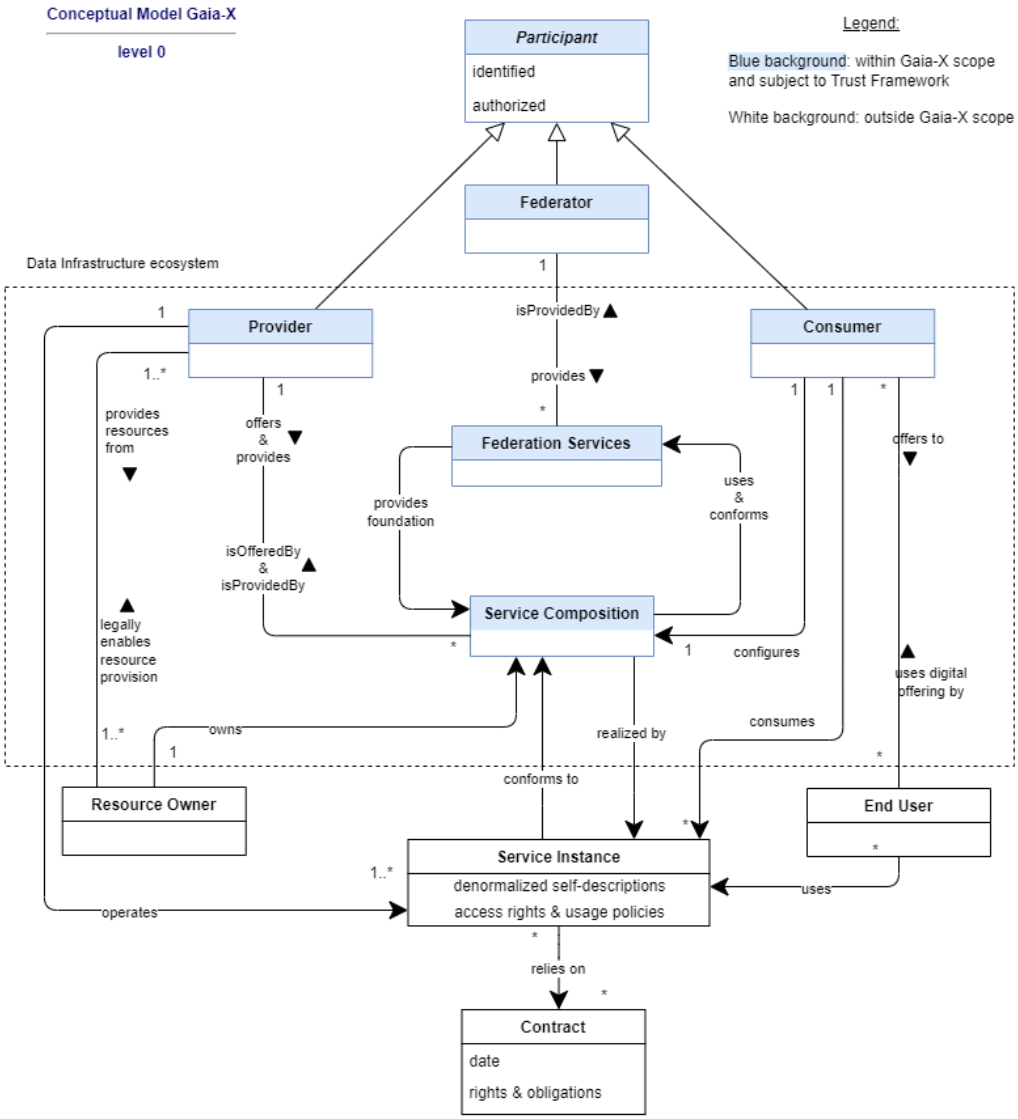


Figure 12 - Gaia-X conceptual model (Gaia-X Association, 2024).

The principal objective is to provide a collaborative environment setting the base for transparency, data sovereignty, and innovation in the digital economy. It involves various kinds

of participants, Providers, Federators, and Consumers, interacting within the ecosystem (Gaia-X Association, 2024).

- **Providers**: They are organizations that offer infrastructure, platform, or software services in the Gaia-X ecosystem. They ensure that their services conform to Gaia-X standards and can integrate seamlessly with other providers.
- **Federators**: Federators are responsible for managing federated services and ensure that each of the participants in it conforms to the standards of Gaia-X. Federators are an integral part of governance, compliance, and generally, orchestration of services within an ecosystem.

- **Consumers**: Any individual or legal entity using the services offered in the Gaia-X framework. Consumers have gained interoperability and flexibility via Gaia-X-compliant services.

3.5.1 Service Composition Model

The Gaia-X service composition model addresses the dynamic composition and management of services within the ecosystem. It supports the creation of composite services that originate from different providers. This model provides flexibility, scalability, and interoperability with a focus on standardized interfaces and APIs for smooth integration and management of services. It provides the mechanisms for service discovery, orchestration, and lifecycle management, which enable users to further customize and optimize their service composition to comply with some specific requirements while remaining standards-compliant with Gaia-X. (Gaia-X Association, 2024).

The figure 13 illustrates the service composition model.

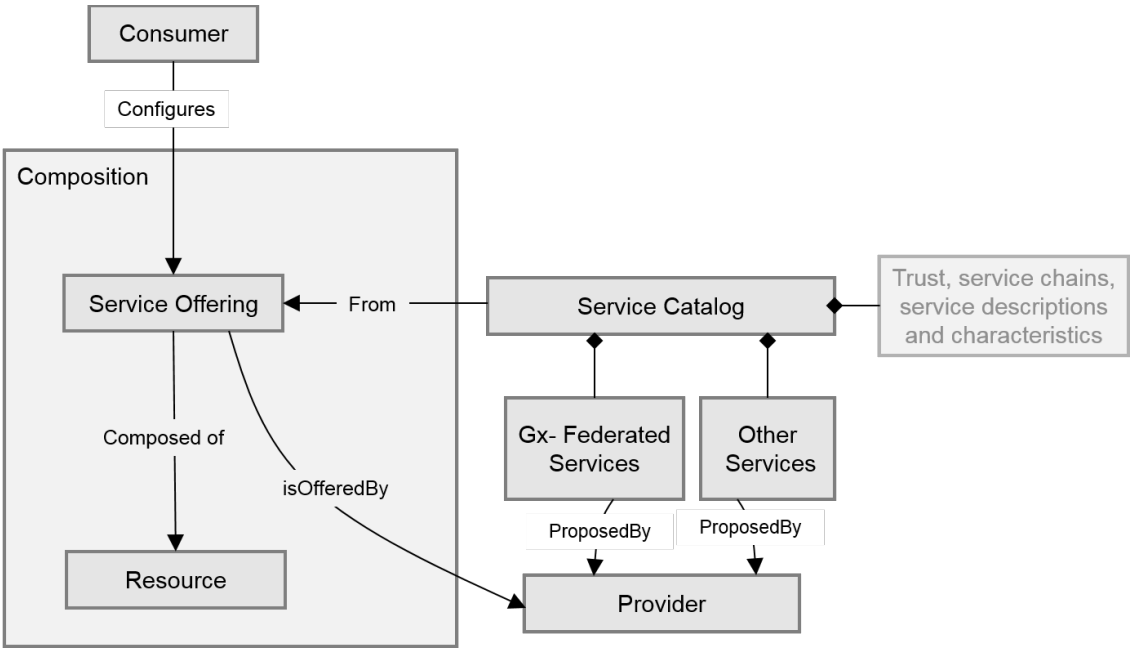


Figure 13 - Simplified and abstract conceptual service composition model (Gaia-X Association, 2024).

Resources within Gaia-X are differentiated into: (Gaia-X Association, 2024).

- **Physical Resources**: These are tangible resources, such as servers, storage devices, and networking equipment, that provide an infrastructure backbone for Gaia-X.

- **Virtual Resources**: This is a resource abstraction, which offers service or platform features at a higher level to the end-user for data processing, analytics, and application hosting.

- **Instantiated Resources**: These are configured instances of virtual resources utilized by consumers to achieve specific tasks or applications.

3.5.2 Interconnection Point Identifier (ICP)

ICPs form the core of the Gaia-X architecture and design, acting as standardized access points by which various services and data will be exchanged between providers. They ensure that high efficiency and seamless interoperability will characterize the ecosystem's operations. (Gaia-X Association, 2024).

Table 2 – ICP Key Features (Gaia-X Association, 2024).

Key Features	Rationale
Standardized Identification	Provides a standardized system of identification for interconnection points over the network so that services provided by different providers can be connected coherently and effectively.
Routing and Access Management	Optimizes how information and service requests are routed. In this respect, they ensure that the flow of data is directly and efficiently provided from the providers to consumers. They improve access control, reduce latency times, and improve response times.
Scalability and flexibility	Standardization of the interconnectivity allows the ICP to enable rapid scaling of the ecosystem with easy onboarding of new services and providers and with minimal disruptions. Flexibility also permits, considering the evolution of emerging user needs, access to a wide range of services from various providers.
Interoperability	ICPs plays a major role in achieving interoperability within the Gaia-X ecosystem, whereby different systems and services shall be able to interoperate irrespective of their native technologies or platforms. This can also mean the combination of services provided by several parties into more complex offerings for seamless user experience.
Security and Compliance	ICPs give security to the data that will be transmitted between the interconnection points, following very strict standards for security and compliance provided by Gaia-X. They ensure integrity and data privacy in such a way that all exchanges become compliant with regulations and meet user expectations.
Network Integration	ICPs now allow integration of hybrid cloud environments and make the management of public or private networks possible. This enables the usage of various cloud offerings while keeping full control over data and services. Furthermore, ICPs support the connectivity between legacy systems and modern cloud services, thus ensuring seamless transitions and upgrade paths within the ecosystem.

3.6 Policies

The Gaia-X Provider Policy stipulates that providers shall be responsible for the security, interoperability, and sovereignty of the provided services in a harmonized manner. The providers must comply with the regulations of the GDPR and be certified for conformance to Gaia-X. The usage policies define the use of the services concerning transparency and user consent (Gaia-X Association, 2024).

The Consumer Policy stresses the protection of the right of users to control their data through portability and privacy, hence the movement of data from one provider to another with no threat to compliance. All these policies put together are aimed at building a trustworthy, efficient digital ecosystem (Gaia-X Association, 2024).

3.7 Service Offering, Service instances and services contracts

In Gaia-X, service offerings denote all the various services offered by providers—infrastructures, platforms, and software. These offers are characterized by capabilities, service levels, prices, and compliance with Gaia-X standards, so consumers can select according to their individual needs. Providers characterize their services through standardized metadata for transparency and interoperability, thus enabling consumers to compare and select the most appropriate services within the ecosystem (Gaia-X Association, 2024).

Service instances are concrete deployments of such offerings, which are configured according to consumer requirements. They are dynamic, scalable, and perfectly integrated with other services while being interoperable with the ecosystem. Service contracts will formalize these agreements between providers and consumers, including terms for service levels, security, and compliance obligations. Therefore, these contracts will form the basis of trust and accountability by using computable contracts to automate compliance checks and service level monitoring to realize reliability and efficiency in managing digital services within Gaia-X (Gaia-X Association, 2024).

3.8 Contract

Contracts are important for Gaia-X because they draw a very fine line on what the service provider is supposed to deliver to the consumers based on agreed terms and conditions. They detail the terms of service at detailed levels regarding service levels, security, and compliance that introduce accountability and transparency. They spell out performance metrics and data protection measures, among other critical terms that should ultimately guide service delivery and usage (Gaia-X Association, 2024).

Another core innovation of Gaia-X is Computable Contracts as a Service. They digitize the contracts between parties and automate compliance checks with service level monitoring for

increased reliability and efficiency. Computational contracts arise from the terms and conditions of traditional contracts, translated into a format readable by machines, thereby allowing not only execution but also automatic enforcement. This automation reduces administrative overhead and guarantees consistent application of the terms of the contract, hence offering a clear and streamlined way of managing service contracts or agreements.

Computable contracts would mean that services are dynamically adjusted, therefore adjusted in real-time as often as required because of changes in requirements or needs of compliance. They would provide a means for providers and consumers to manage efficiently their respective obligations and rights and hence foster a more adaptive and responsive ecosystem. Gaia-X also builds on computable contracts and advanced technologies such as smart contracts and blockchain to build a robust framework that can support secure and trusted digital service delivery in a federated environment (Gaia-X Association, 2024).

3.9 Self-Descriptions

Self-descriptions are at the heart of the Gaia-X framework because they secure transparency, interoperability, and trust in a federated data ecosystem. Basically, self-descriptions are structured digital representations containing overall information on entities—for instance, service providers, services, and data assets. They enable efficient discovery, rating, and integration of services by the standardized information they convey, which can be automatically processed within the Gaia-X ecosystem (Gaia-X Association, 2024).

3.9.1 Definition of Self-Description:

In Gaia-X, a Self-Description is an official description given by an entity of its attributes, capabilities, and compliance status. It is comparable to some sort of digital profile that communicates the core characteristics of an entity to other participants in the ecosystem. A description of this nature is critical to building trust and being interoperable by allowing entities to clearly announce their capabilities and confirm adherence to Gaia-X standards (Gaia-X Association, 2024).

3.9.2 Self-Description Structure

A self-description would include a set of metadata elements that will capture information about the identity, functionality, and conformance of an entity. Of course, such self-descriptions will always have the following main components: (Gaia-X Association, 2024)

- **Identity Information**: It includes details about the entity's identity, which is its name, type, and unique identifiers.

- **Capabilities**: Services or functionalities offered by the entity along with technical specifications and performance metrics.

- **Compliance and Certification**: Information about the entity's compliance with applicable standards and regulations, and related certifications.

- **Security Features**: Information on security features implemented by the entity concerning data protection and the safety of operations.

3.9.3 Self-Description Schema

The self-description schema defines the structure and semantics for constructing self-descriptions, hence creating coherence and interoperability throughout the Gaia-X ecosystem. The schema constrains the data elements to be provided, their relationships, and related constraints to let an entity precisely manifest its attributes. The schema is extensible, meaning it allows an entity to extend its self-descriptions with information that is important in its context without deviating from the Gaia-X standards.

3.9.4 Self-Description Life Cycle

The following steps are involved in the life cycle for a self-description (Gaia-X Association, 2024).

- **Generation**: This step involves the gathering of all kinds of relevant information that concerns the entity and its encoding into standardized formats of self-descriptions.

- **Validation**: The created self-description will be checked against the Gaia-X schema to prove standards compliance and correctness.

- **Registration**: The valid self-description will be registered within the Gaia-X ecosystem, and thus it will be available to other participants for discovery and assessment.

- **Update and Maintenance**: Since the entity is subject to change, so must its self-description; for example, in case of changes to the services offered, capabilities, or compliance status. This continuous process is necessary to maintain correctness and relevance.

- **De-commissioning**: In case an entity does not exist anymore or is no longer willing to belong to the Gaia-X ecosystem, de-commissioning of its Self-Description removes it from active use and thus keeps the Ecosystem up to date.

3.9.5 Self-Description Creation

Self-descriptions are key components of Gaia-X, as they define the identity and functionalities of some entities, such as service providers, services, or data assets. This calls for the representation of all Gaia-X participants in such a way that they are precisely discoverable and evaluable against standardized information (Gaia-X Association, 2024).

Basically, the process for creating self-descriptions consists of the following important steps:

Claims Collection - First, entities gather claims with respect to their attributes, capabilities, and compliance status. Claims are factual assertions describing what the offering of an entity is and how an entity works. This includes technical documentation, compliance reports, and operational data (Gaia-X Association, 2024).

The figure 15 illustrate the process of collecting signed claims.

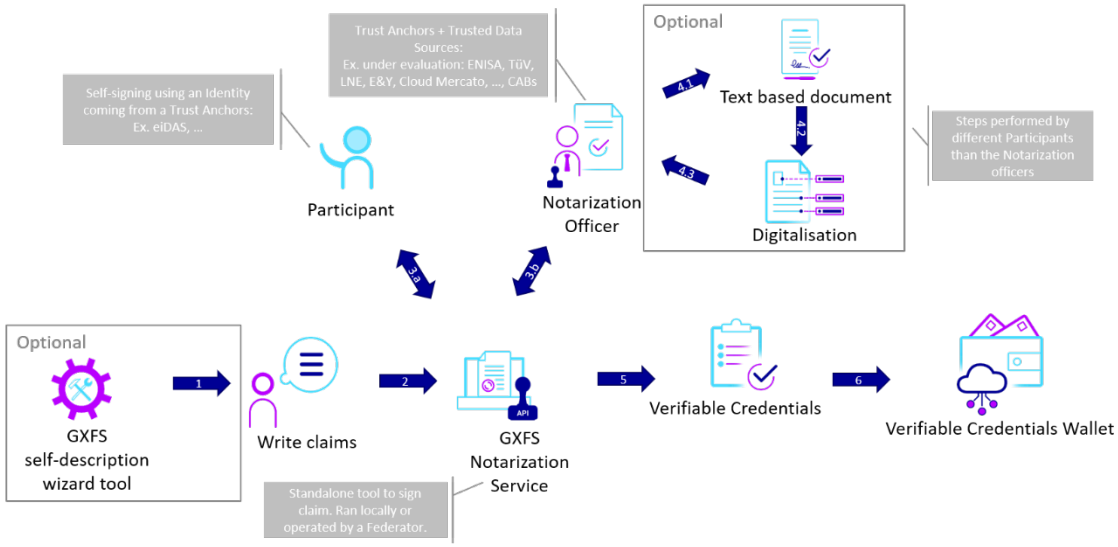


Figure 14 - Collecting signed claims (Gaia-X Association, 2024).

Gaia-X Verification – This would be followed by verification of the self-description for accuracy and correctness against the standard of Gaia-X. This may include automatic checks to validate data against the Gaia-X schema and check whether relevant fields are present and correctly formatted. Gaia-X verification can include third-party audits or peer reviews to build more trust and credibility in the information provided (Gaia-X Association, 2024).

The figure 16 illustrates the Gaia-X Verification.

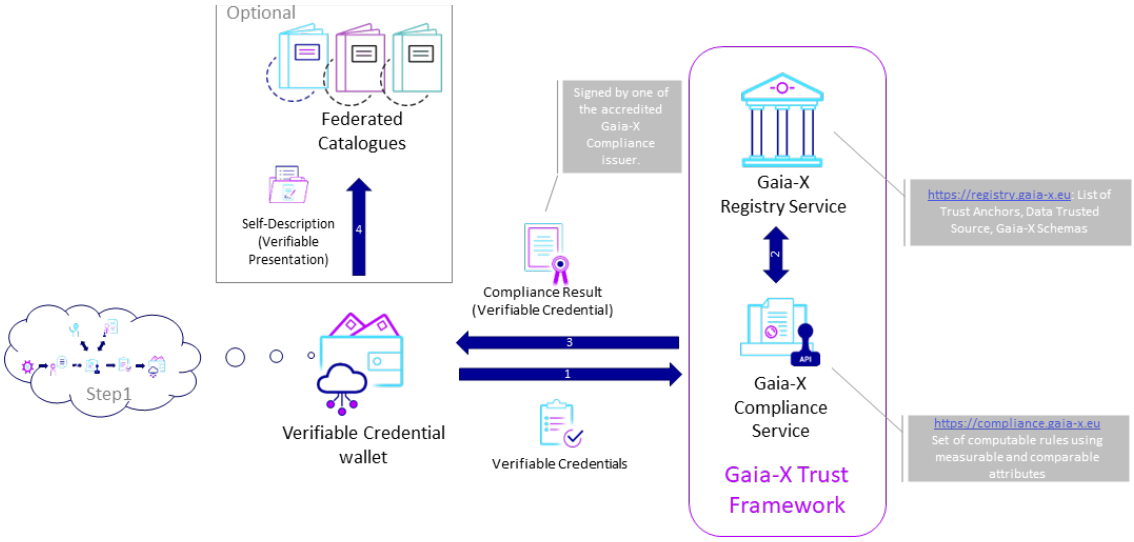


Figure 15 - Validating signed Claims using the Gaia-X Trust Framework (Gaia-X Association, 2024).

Federation Governance - It uses the same workflow as the Gaia-X verification but where it differs is that every federation is free to extend Gaia-X governance with the possibility of adding custom rules and checks.

The figure 17 represent federation extending Gaia-X governance.

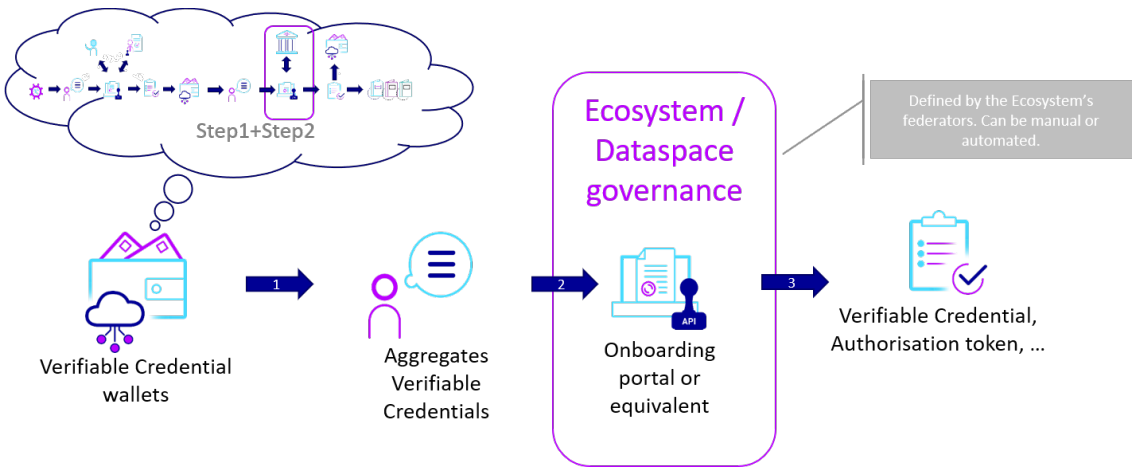


Figure 16 - Federation extending Gaia-X governance (Gaia-X Association, 2024).

3.10 Gaia-X Operation Model

The operational model of Gaia-X describes the organizational structures, processes, and governance mechanisms that would ensure effective maintenance and management of the Gaia-X ecosystem. It is focused on enabling a federated data infrastructure while fostering data sovereignty, security, and interoperability on and beyond Europe (Gaia-X Association, 2024).

The first part of this chapter introduces the Gaia-X Ecosystem, as well as Trust Anchors. Trust Anchors are defined, including details about who defines them and how they will be nominated.

The second part defines Gaia-X Compliance, and how to become compliant. It introduces the Gaia-X Compliance Service as well as the usage of Gaia-X Labels.

Finally, the last section will cover the Gaia-X Self-Descriptions life cycle and the Gaia-X Registry, which provides essential support for the Gaia-X Decentralized Autonomous Ecosystem.

3.10.1 Trust anchors

A trust anchor is an institutional or other entity or mechanism in Gaia-X that can constitute a source of trust in the network. In this respect, they identify and verify the identity of participants, attesting to the achievement of certain standards or criteria. In doing so, they establish the relevant trust relationships between actors within an ecosystem, be they service providers, consumers, or authorities. The trust anchors ensure the integrity and reliability of the ecosystem by ensuring all participants play by a common set of principles and standards (Gaia-X Association, 2024).

3.10.2 Gaia-X Labels

Gaia-X Labels are certifications or marks of compliance that a participant adheres to the standards and principles laid down by Gaia-X. These labels enable any easy reference by users or organizations for services and providers adhering to specific requirements set by the Gaia-X standards concerning data security, interoperability, and transparency. The Gaia-X Labels are indicative of quality and engender trust in the user by giving transparency over the assurance of compliance and conformity—facilitating selection by the user within the ecosystem (Gaia-X Association, 2024).

3.10.3 Gaia-X self-description

Gaia-X Self-Description is how participants can provide a description of their services, capabilities, and standards compliance at Gaia-X in structured and machine-readable forms.

It would therefore indicate metadata for services offered by the participant, including technical specifications applied, certifications, and data protection conformance. In this respect, self-

description enables automated processes to support service discovery and valuation within the Gaia-X ecosystem (Gaia-X Association, 2024).

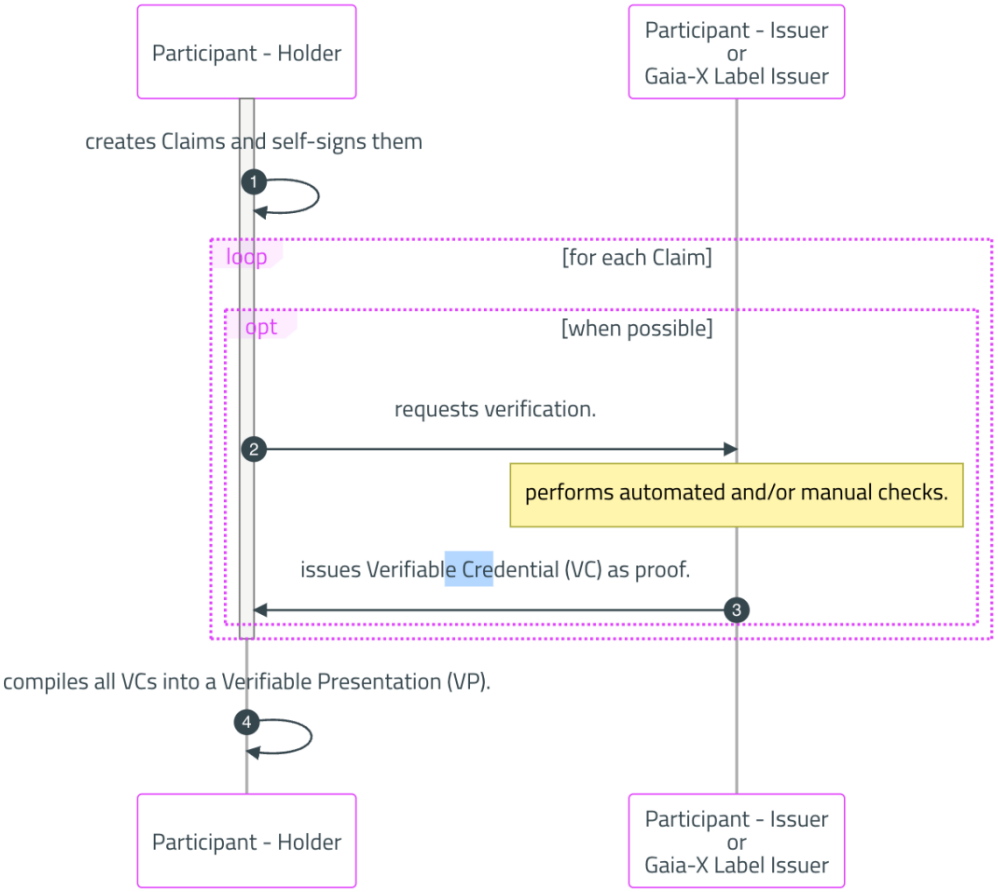


Figure 17 - Compiles all VCs into a Verifiable Presentation (VP) (Gaia-X Association, 2024).

The figure 18 shows the process of creating a self-description where the participant creates a claim and for each claim created sends to a Gaia-X Label issuer and there after it is issue a verifiable credential as proof.

Gaia-X Self Descriptions are created using the following vocabulary of the W3C Verifiable Credentials Data Model standard.

W3C Term	Example with a car
Claim	My car is red
Verifiable Credential	The garage's attestation that my car is red
Verifiable Presentation	Me showing to my friend the garage's attestation that my car is red
Issuer	The garage
Holder	Myself
Verifier	My friends

Figure 18 - Gaia-X Self-Description Vocabulary (Gaia-X Association, 2024).

3.10.4 Difference Between Self-Description Proofs (VC), Gaia-X Trust Framework, and Gaia-X Labels

VC Self-Description Proofs: Verifiable Credentials used by each party to establish the validity of their self-descriptions, thereby ensuring that information provided in self-description is correct and reliable.

Gaia-X Trust Framework: The overarching set of rules and standards controlling trust within the ecosystem. It would ensure that all actors will adhere to a common baseline of principles and thus help in maintaining a secure and reliable environment.

Gaia-X Labels are certificates that attest to each participant's conformance to the Gaia-X standards. This means, once it sets the standards, the Trust Framework really would, and the labels serve as visible evidence of compliance.

3.10.5 Self-description compliance

A Self-Description must be subject to an instance of the Gaia-X Compliance Service to qualify it as Gaia-X compliant. At submission time, results on compliance are captured in two ways:

1. If compliance is validated, the service issues a Verifiable Credential which can be enclosed into the SelfDescription, in line with the self-sovereign principle where the holder controls the information.

2. If the Gaia-X Compliance Service is accessed via the Gaia-X Registry, the registry is going to emit an event for the synchronization of Catalogues with the URL of the Self-Description file. The next section further elaborates on the specifics of the Gaia-X Registry. (Gaia-X Association, 2024c)

The Figure 19 illustrates the self-description compliance.

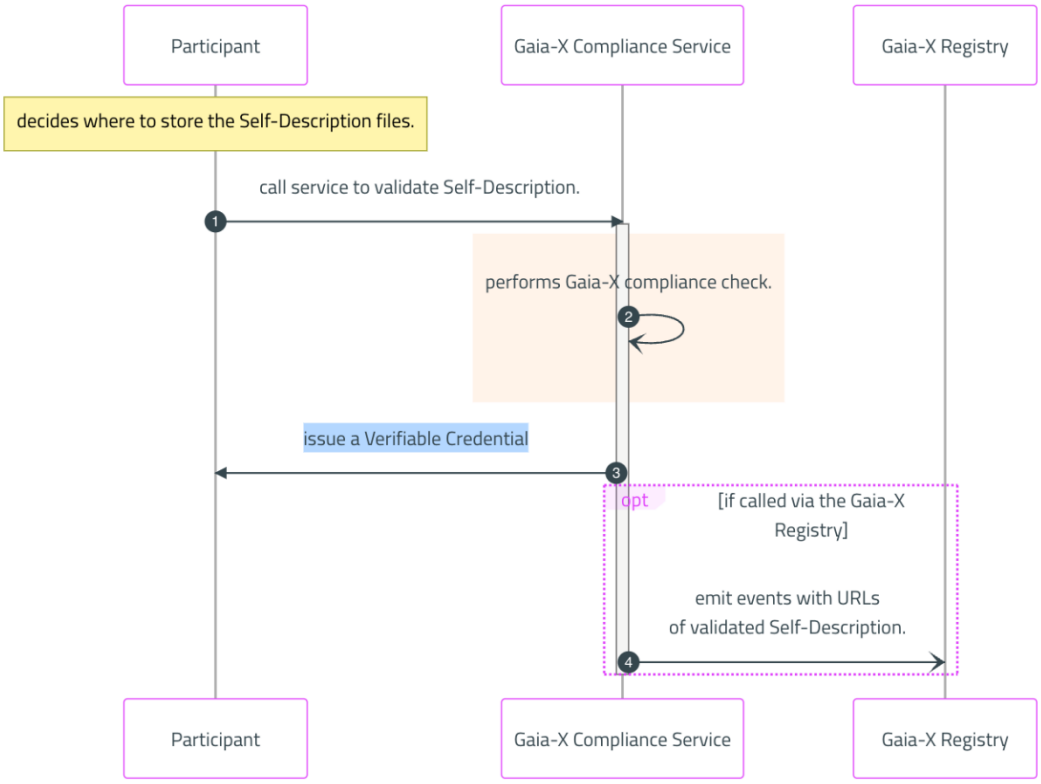


Figure 19 - Self-Description Compliance (Gaia-X Association, 2024c)

3.10.6 Self-description remediation

Self-Descriptions can also expire over time and be basically stamped as being in one of three states:

- **Expired**, after a timeout date, for example, expiry of a cryptographic signature.
- **Deprecated**, when replaced with a new Self-Description.
- **Revoked**, by its original issuer or any other trusted party owing to incorrect or fraudulent information.

Even if the Expired and Deprecated states of the Self-Description may be automatically created, based on information stored in the Gaia-X Registry or Catalogues without an extra process, the next paragraphs describe how Self-Descriptions are revoked. The operational model will discourage malicious actors from intentionally spoiling the Registry and Catalogues since Gaia-X compliance will come to progressively cover principles like interoperability, portability, and security.

Revocation of Self-Descriptions can occur in several ways: by authorship, where the author explicitly revokes or deprecates it; by automation, where the Gaia-X Compliance Service discovers non-compliant attributes; and by manual decision after an audit by a Gaia-X Participant. In case some attribute is found incorrect, suspension is automatically done, and the revocation is forwarded to the Gaia-X Association for approval, thus providing an opportunity to the owner of the Self-Description to present his case in days. The results from voting shall therefore be made visible and stored in the Gaia-X Registry or any local Ecosystem's Registry to help reduce subjective decisions and increase transparency (Gaia-X Association, 2024c).

As we can see, the entire structure of the Gaia-X ecosystem marks a new frontier in the digital infrastructure space for ensuring data sovereignty, security, and interoperability. Besides, it creates a truly federated environment wherein seamless data exchange and thereby enabling collaboration across industries is realized under European regulations. This chapter has outlined the importance of Federation Services, compliance mechanisms, and data exchange as cornerstones of the Gaia-X framework. These will help build up the trust but at the same time create a fuller, more scalable, and secure ecosystem to help a variety of sectors, including energy to healthcare.

This means dealing with the integration of SSI into how Gaia-X has so far been built. SSI is crucial in extending user control and privacy in the platform offered by Gaia-X because it allows individuals and organizations to become their own custodians over their digital identity. The role of SSI at Gaia-X is discussed in the section below. It will outline why SSI forms an integral part of establishing trust and compliance around decentralized data sharing.

3.11 SSI in Gaia-X

3.11.1 Introduction

Today, we are most often reduced to using a single centralized cloud identity provider with all its pros and cons. However, due to the digitization of the world in the past couple of decades and a shift in people's mindsets, a range of new decentralized Web3 components is fast taking center stage. Those innovations aspire to return the decentralized nature to the Internet, as it was originally intended. One of these thoughts is self-sovereign identity, founded on the principle of Web of Trust, which empowers a person again with the possibility of regaining control over his or her own identity and data (Maier and Pohlmann, n.d.).

3.11.2 Basic structure and process of the SSI ecosystem

At its very core, SSI works on a very decentralized model where users are in direct control of their digital identity and verifiable credentials, independent of centralized authorities. In such an ecosystem, there are three main roles interacting with each other according to the "trust triangle" model: an issuer who creates verifiable digital credentials including a given claim, such as certificates or authorizations; a holder who safely manages those credentials within an SSI wallet and selects which credential is to be shared and with whom; and a verifier who requests and checks credentials to process and validate certain claims securely in a peer-to-peer connection. A decentralized approach removes the need for any middlemen and ensures that all data belongs to and is controlled by users, increasing privacy, security, and trust in all digital interactions (Maier and Pohlmann, n.d.).

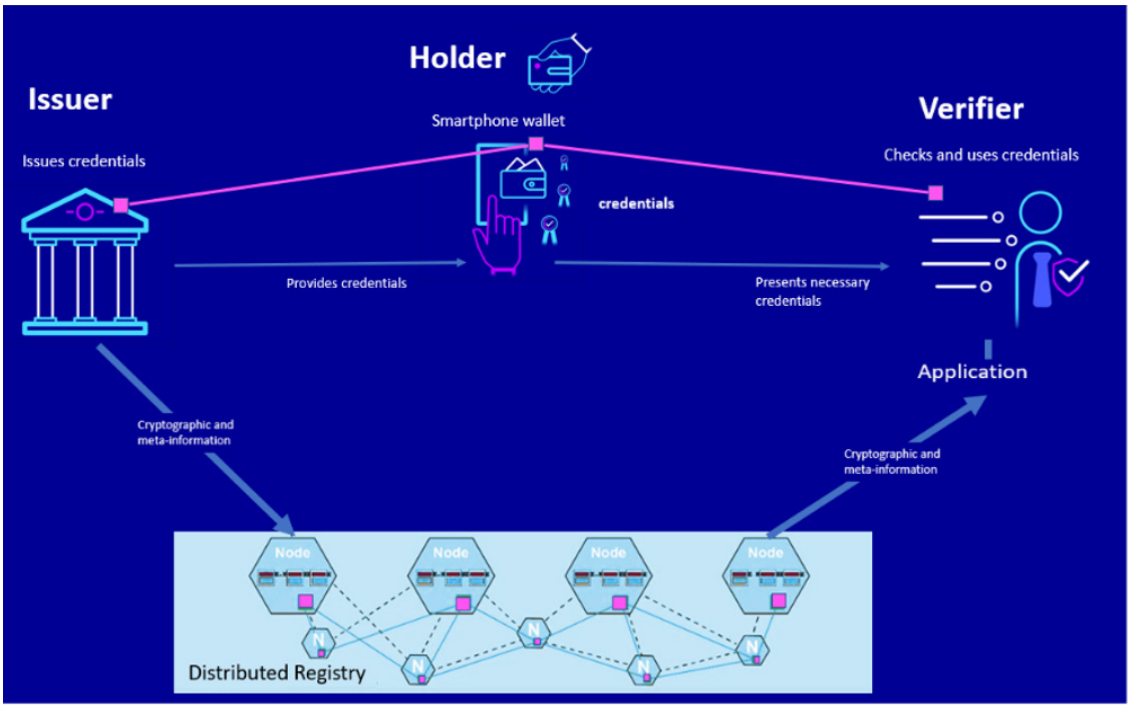


Figure 20 – SSI Ecosystem for digital identities and credentials (Maier and Pohlmann, n.d.)

3.11.3 Use of SSI in Gaia-X Federation Services

SSI is, therefore, an integral part of Gaia-X Federation Services in developing a decentralized, secure digital ecosystem. Unlike the conventional centralized systems, wherein one authority manages identities and credentials, Gaia-X utilizes SSI to help participants manage their digital identities and credentials autonomously. This type of decentralization makes it easier for participants in Gaia-X to be fully in charge of their data, hence offering both privacy and trust. The SSI framework of Gaia-X allows for decentralized authentication via SIOP, attribute-based authorization by credentials, and self-management of digital identity through secure wallets and agents. These enable participants to communicate directly and securely amongst each other, without the need for any central intermediary, and hence foster a more open and trusting digital ecosystem (Maier and Pohlmann, n.d.).

In addition, federation services can be enabled in Gaia-X by SSI, ensuring trustworthiness in data exchange and compliance with privacy standards. The Gaia-X Federation Services, for example, leverage SSI to deliver verifiable credentials at the heart of trust frameworks and labeling concepts that ensure conformance of services to exacting trust and compliance requirements. Moreover, SSI supports selective disclosure of data and advances in data sovereignty—users only need to disclose the least amount of information. It supports a federated catalog system for the generation of trustworthy data for search indexes and secure interlinking of services in SSI-based applications and provides means for the notarization of credentials and management of compliance and fraud controls. In this way, Gaia-X is ensured of the possibility to support a

wide spectrum of applications and services securely and in decentralized form (Maier and Pohlmann, n.d.)

3.11.4 SSI in the Context of Data Exchange and Data Protection

Beyond data encryption, the realization of SSI concepts, such as ZKPs and ABAC, helps to further develop future zero-trust and privacy-preserving architectures that avoid the intrinsic weaknesses of traditional static RBAC. On the other hand, ZKPs will turn into important cryptographic tools, which will make compliance and efficient exchange of verifiable credentials in the frameworks of SSI compliant with data protection (Maier and Pohlmann, n.d.).

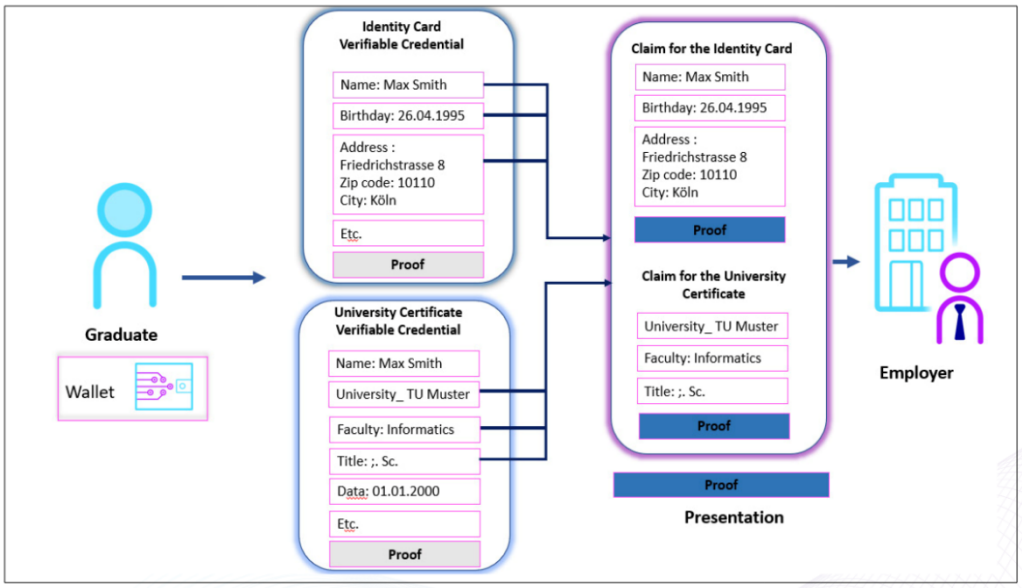


Figure 21 - Overview of SSI Architecture (Maier and Pohlmann, n.d.)

- **Selective Disclosure**, in proving an identity or credential, a user will only share the relevant claims from the VCs, which turn reveal only necessary information. For instance, if an application were to request proof of a user's address, the user could select and disclose only the name and address from the VC of the identity card, without other details like date of birth or height. This approach is vastly different from the physical credential—not, say, a passport—where selective disclosure isn't possible (Maier and Pohlmann, n.d.).

- **Predicate Proofs**, which enable users to prove some facts without necessarily sharing the actual data. For example, instead of disclosing a specific age, a user might prove having passed the age of 18 using a "greater than 18" predicate proof (Maier and Pohlmann, n.d.).

- **Signature Blinding**, which might serve as a unique identifier otherwise, are prevented from being correlated using signature blinding. This approach randomizes the digital

signature before it leaves your server so that the authenticity and integrity of the VCs are maintained, although it is protecting against possible data correlations (Maier and Pohlmann, n.d.).

- One such approach is based on something called **Private Holder Binding**. This would cryptographically bind a Verifiable Credential to a user without exposing a user's DID. In this approach, VCs are indirectly bound with a user through individual link secrets, so there would be no need for direct correlation of the DID (Maier and Pohlmann, n.d.) .

3.12 Limitations to implementing Gaia-X in the energy sector

There is an interesting paper about the limitations to implementing Gaia-X in the energy sector courtesy of (Gaia-X Association, 2021) that one can consider a good read going forward. The paper explains several points:

- **Data Quality and Transmission**: The existing model of the energy market in data exchange is not able to serve future requirements on data transmission, quality, and security. Real-time data with great granularity—like smart meter data—shall be integrated, which becomes quite a challenge under the present system. (Gaia-X Association, 2021)

- I**ntegration of new players and technologies**: No integration for new players such as e-mobility providers and prosumers, and technologies like mass-market heat pumps. These are expected to assume key roles in the energy market of the future but have not yet been integrated into the current system. (Gaia-X Association, 2021)

- **Interoperability and Portability**: The Gaia-X infrastructure shall, in its full effectiveness, support interoperability and portability of data across these sectors. This would translate to mean the making of an ecosystem where companies and other actors develop new kinds of products, services, and business models with access to this data. (Gaia-X Association, 2021)

- **Data Sovereignty and Security:** Guaranteeing the sovereignty and security of data, either from or to the different actors of the energy market, is one of the main challenges. It says that the documents stress policies are to guarantee a high standard of data security and protect data sovereignty. (Gaia-X Association, 2021)

- **Technical and Organizational Challenges**: Major technical and organizational challenges exist in handling the heterogeneous data to be handled and the complexity that comes therewith in an effective operation of energy markets. This refers to the

requirement for harmonized data exchange protocols between different market players with standardized data semantics. (Gaia-X Association, 2021)

4 Realizing Energy Dataspace

4.1 Introduction

In this chapter, we will examine the Gaia-X Digital Clearing House and the Sovity Data Connector while we try to come up with at least a minimally Gaia-X-compliant dataspace. Using those tools we want to flesh out—in an abstract manner—the primary components and processes necessary to achieve interoperability, data sovereignty, and compliance within a Gaia-X framework.

We have the following diagram for a better understand:

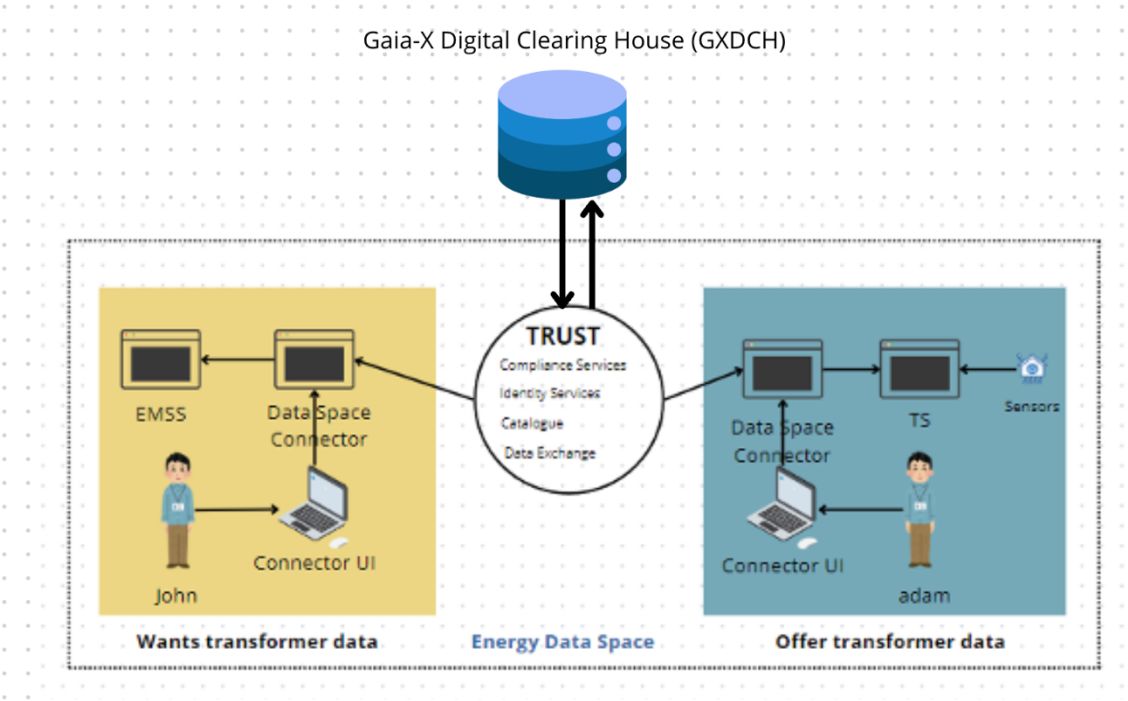


Figure 22 - Energy Data space with Gaia-X digital clearing House

As we can see, the Gaia-X Digital Clearing House (GXDCH) will be responsible for handling the trust, compliance services, identity services and securing the data exchange and for the data space connector we are going to use the sovity data space connector because it already as most of the core components of an data space available out of the box along with a user interface ensuring that we can build a minimum viable data space with ease.

4.2 Gaia-X Digital Clearing House (GXDCH)

According to the website, Gaia-X Digital Clearing House is an enforcement node for Gaia-X rules and makes being Gaia-X compliant easier by onboarding into the ecosystem. These nodes are interchangeable and operated by different market players, acting as Gaia-X Federators who run core services needed for compliance and onboarding new participants. It provides the interface to external Trust Anchors like e.g. Conformance Assessment Bodies but also to identity verification systems like eIDAS and other Trusted Data Sources defined by Gaia-X AISBL. This integration will ensure secure and reliable data exchange within the Gaia-X framework. (Gaia-X Association, 2024d)

In short, GXCDH is the Gaia-X compliance enabler and one entry point to multiple services within the Gaia-X ecosystem. It serves as the single-entry point into Gaia-X for both mandatory and optional services for meeting various requirements.

Core Mandatory Components of GXDCH:

- **Gaia-X Registry**: This component acts more like a directory, keeping stock of participants and services involved in the Gaia-X ecosystem. It makes sure that there is a registry of all the entities and services on the network so that they can be identified by their identities. (Gaia-X Association, 2024d)

- **Gaia-X Compliance**: This ensures that services and participants conform to the standards of Gaia-X, ensuring secure interactions and compliance with rules. (Gaia-X Association, 2024d)

- **Gaia-X Notarization Service**: This service provides a registration number to the participants. Somehow, this number behaves like a validation such that all the entities in Gaia-X are then authenticated and trusted. (Gaia-X Association, 2024d)

Optional Components of GXDCH:

- **Wizard**: This component facilitates easy setup and management of a user's services within the Gaia-X ecosystem. It facilitates integration and operational tasks more easily. (Gaia-X Association, 2024d)

- **Wallet**: Does the job of managing credentials and identities thus enabling the safe storing and processing of a user's digital identities along with associated data based on the concept of Gaia-X. (Gaia-X Association, 2024d)

- **Catalogue**: The Catalogue brings structure to the large variety of Gaia-X services and enables users to find services and use them within the ecosystem very easily. (Gaia-X Association, 2024d)

4.2.1.1 Installing and testing

The GXDCH components may be installed locally for testing. Typically, installation involves creating Docker containers or Kubernetes clusters for a component depending on the component type; in most cases, environment variables must be defined for each component.

As the compliance service tries to contact the registry, it is advised to deploy the components in the following order:

1. **Registry**
2. **Compliance**
3. **Notary**

Now we are going to see how to deploy locally each of the components.

Gaia-X Lab Compliance Service

In the Gaia-X repository, setting up the service consists of 3 steps:

Step 1: Generating a certificate

```
PS C:\Users\ricardo> $ openssl req -nodes -x509 -newkey rsa:2048 -keyout key.pem -out cert.pem -sha256 -days 365
```

The command above will generate 2 files: cert.pem for the certificate and key.pem for the private key.

Step 2: Setting up the compliance service

We need to perform the following commands:

- `git clone https://gitlab.com/Gaia-X/lab/compliance/gx-compliance.git`
- `cd gx-compliance`
- `nvm install`
- `npm install`

After performing a npm install we need to set up the environment variables in the root directory of the project.

```
You, 4 months ago | 2 authors (You and one other)
X509_CERTIFICATE=`-----BEGIN CERTIFICATE-----
copy `cert.pem` content
-----END CERTIFICATE-----`
privateKey=`-----BEGIN PRIVATE KEY-----
copy `key.pem` content
-----END PRIVATE KEY-----`
privateKeyAlg=`PS256`
REGISTRY_URL='https://registry.lab.gaia-x.eu/development'
BASE_URL='https://localhost:3000'
NODE_TLS_REJECT_UNAUTHORIZED='0'
LOCAL_HTTPS='true'
DISABLE_SIGNATURE_CHECK='true'
```

Figure 23 - Configuration File for **Gaia-X Lab Compliance Service**

Then copy the certificate cert.pem into gx-compliance/src/static/.well-known/x509CertificateChain.pem and replace the existing certificate chain with the generated cert.pem. It is advised to run npm run build after every change to BASE_URL or x509CertificateChain.pem. And finally, we can start the compliance service by running either npm run start and npm run start:dev.

Step 3: Sign your VerifiableCredentials

If you have already signed your VC, skip to the end. With key.pem and cert.pm, you are ready to sign your VC. You can do this manually or by using the Lab wizard (https://wizard.lab.Gaia-X.eu/) . For doing it manually, you can use library Gaia-X JsonWebSignature2020. Once your VC is signed, go to the compliance service https://localhost:3000/docs/#/credential-offer/CommonController_issueVC and ask to have your VerifiablePresentation signed, resulting in a compliance VerifiableCredential.

The figure 24 describes the creating of a Verificable Credentials in two ways, manually where the user prepares the vc shapes and signs it or by using the gx-signing/Gaia-X wizard (figure 25) to do so.

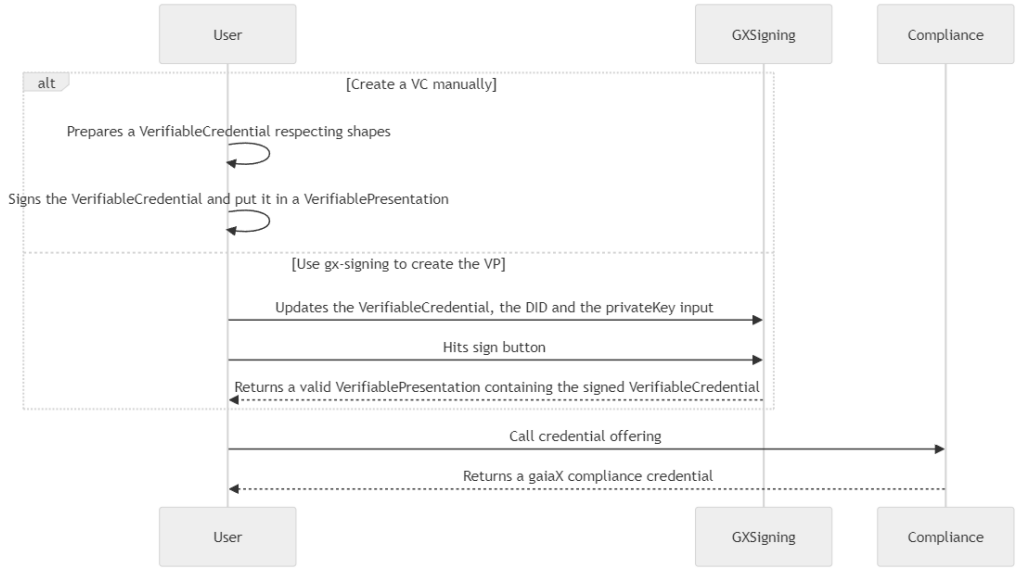


Figure 24 - Gaia-X compliance workflow (Gaia-X Association, 2024e)

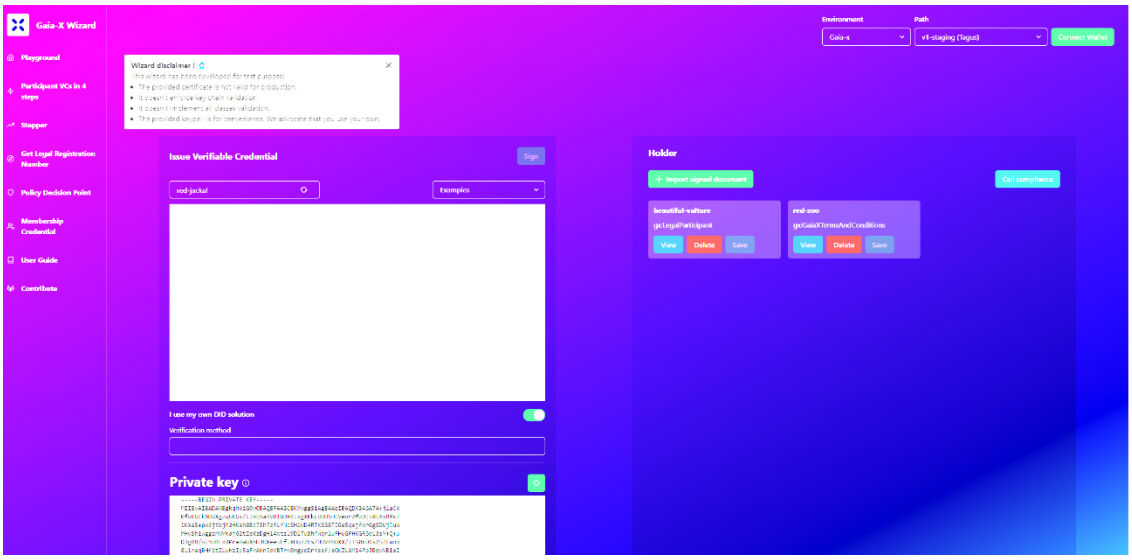


Figure 25 - Gaia-X wizard UI (Gaia-X Association, 2024f)

Gaia-X Registry

The setup process is straightforward where first we need to do a git clone of the repository and execute the following commands in the root of the project:

- `npm install`
- `mkdir ./dist`
- `mv .env.example .env`
- `npm install -g npx`

And if everything is setup correctly execute the command

- docker-compose up

Gaia-X RegistryNumber notarization API

For the setup process we need to do a git clone like before of the repository and then execute the following commands:

- `nvm install`
- `npm install`
- `npm run setup-env`
- `npm run start:dev`

If everything is done correctly, we should have all the 3 instances of the services running on our local machine.

Within the scope of this work, it became clear when trying to implement the demonstration of the Gaia-X Digital Clearing House-GXDCH that the documentation given is currently inadequate, and the technical requirements are very high. These conditions made it difficult to develop a functional implementation in the context of this demonstration. Because of this, the GXDCH could not be integrated or tested in depth. It points out the limitation in terms of more extensive documentation and supporting work that needs to be done in the future to reduce the technical barrier for implementations. Further discussion of these challenges, and critical analysis with potential solutions and recommendations to these challenges, will be made in the Critical Analysis chapter.

4.2.2 Sovity Dataspace Connector

Sovity's CaaS redefines the efficiency, security, and usability of joining and leveraging Data Spaces. Constructed on top of Eclipse Dataspace Components and boosted by experience from Sovity, this CaaS enables companies to quickly connect to the Data Spaces and thus supports compliance with international standards such as IDS and Gaia-X. The service supports the seamless sharing and integration of data, with plans scaling from a basic Starter package to enterprise level, providing flexibility for changing business requirements (Sovity GmbH, 2024).

It is available as a docker image for easy deployment and comes with UI.

Here are some key features:

- **Lightning-Fast Onboarding**: Be connected to your selected Data Space in just minutes— to an intuitive UI at the front end and robust backend.

- **Open-Source Foundation**: It is based on open-source software and thus open and subject to constant improvement.

- **Scalability**: From small business to large enterprise, the CaaS scales with you—from basic connectivity to advanced, enterprise-level data management.

- **Compliance and Certification**: Catena-X certified, IDS and Gaia-X compliant—safe and compliant data exchange is guaranteed.

- **Comprehensive Support**: From each of the various plans, tailored support levels will be found, ranging from self-service resources and professional guidance to 24/7 enterprise support.

4.2.2.1 Quick Setup

To start using sovity EDC first we need to clone the repository and perform the command docker compose up and if everything goes well it should start two local EDC connectors with the following credentials (Sovity, 2024):

Table 3 - Deployment data (Sovity, 2024)

	First Connector	**Second Connector**
Homepage	http://localhost:11000	http://localhost:22000
Management Endpoint	http://localhost:11002/api/management	http://localhost:22002/api/management
Management API Key	ApiKeyDefaultValue	ApiKeyDefaultValue
Connector Endpoint	http://edc:11003/api/dsp Requires Docker Compose Network	http://edc2:11003/api/dsp Requires Docker Compose Network

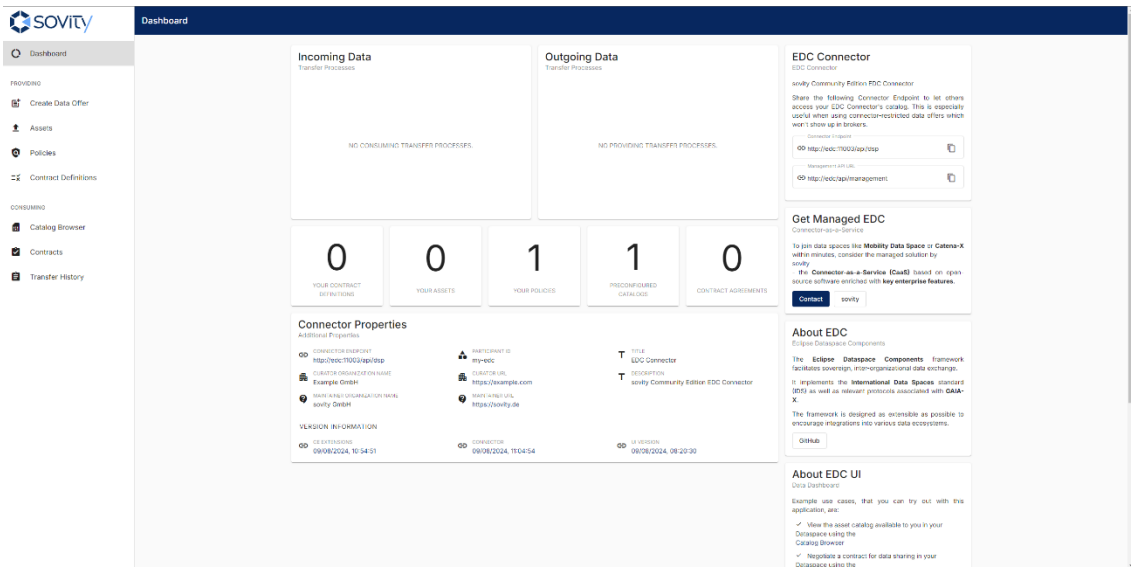


Figure 26 - Sovity EDC Community Edition

For the use case scenario, we are going to be using the sovity Community Edition EDC and it's important to note the limitations (Raghad Matar and Philipp Neuschwander, 2024)

- At present, there are very few policy classes supported by the Sovity Community Edition EDC. It does not support any configuration regarding control of purpose concerning data usage, number of uses, location of consuming connectors, enablement of payments, restrictions on sharing data with other parties, and obligations like deleting the data after a certain period. All these policies need additional customization and development in the connector.

- Most of the technical policy enforcement takes place before transfer. Thus, after the data has been transferred to the consumer's data sink, the Sovity Community Edition EDC provides no technical enforcement. Consequently, the assurance of compliance with a data provider's policies lies solely within organizational measures, for example, legal contracts.

- In the Sovity Community Edition EDC, there are limited possibilities for negotiation. A consumer can only accept all conditions given in an offer, which basically means to take what is given or leave it, whereby consumers cannot make any counteroffer on the same. This may be viewed as unfavorable by those data providers who would like to have more flexibility and control over the terms of contracts.

- The connector automates negotiation, so no manual effort on the part of the data provider is needed to create a contract between two parties. This also means that data providers can't manually inspect contract negotiation requests before deciding to accept them, which in some scenarios—where intricate checking within a contract is necessary—may be considered a disadvantage.

- Sovity Community Edition EDC only supports one of the data transfer methods defined by the EDC framework—not a very serious limitation, considering that this is by far the most used method anyway: HTTPData. This is a method where the provider connector pushes data to an HTTP endpoint specified by the consumer.

4.2.3 Energy data space use case

To realize this use case scenario, we will base our approach on the article (Raghad Matar and Philipp Neuschwander, 2024) where we have two companies involved as follows:

- **An Energy management service system provider (EMSS):** Is a basic service where customers can use to manage their energy consumption from various sources and with that information, it optimizes energy distribution, forecasts demand and improves general energy efficiency.

- **A Transformer service provider:** The TSP specializes in monitoring and managing transformer health and performance. It provides detailed data on transformer load capacity, operational status, and faults that may potentially develop. It finds difficulties in seeking customers and interoperable data exchange implementation. On its part, TSP is concerned with the loss of sovereignty over its data, fearing that information could be utilized for unintended uses or made known to third parties with less than proper consent or compensation.

The **EMSS** would like to buy data on the performance of transformers to enable the optimization of energy distribution, while on the other side, the Transformer Service Provider is interested in selling its transformer data under its own conditions to someone. This is where energy data spaces come into play. In a common data space, both **EMSS** and TSP can easily locate each other. The model shall reduce efforts for collaboration and secure data and service exchange in a trusted environment while guaranteeing that both parties will always keep control over their data but benefit from streamlined, interoperable data exchanges.

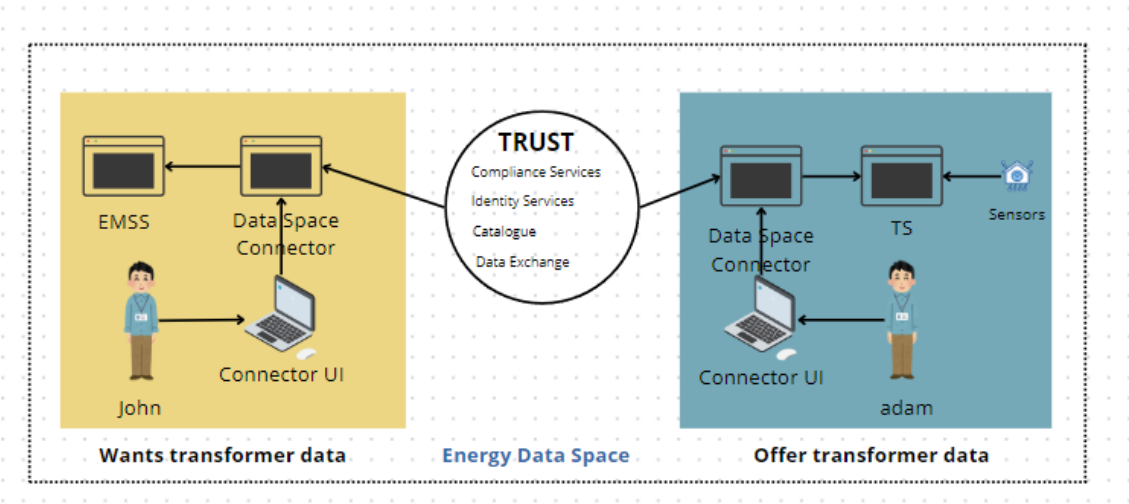


Figure 27 - Energy data space demo

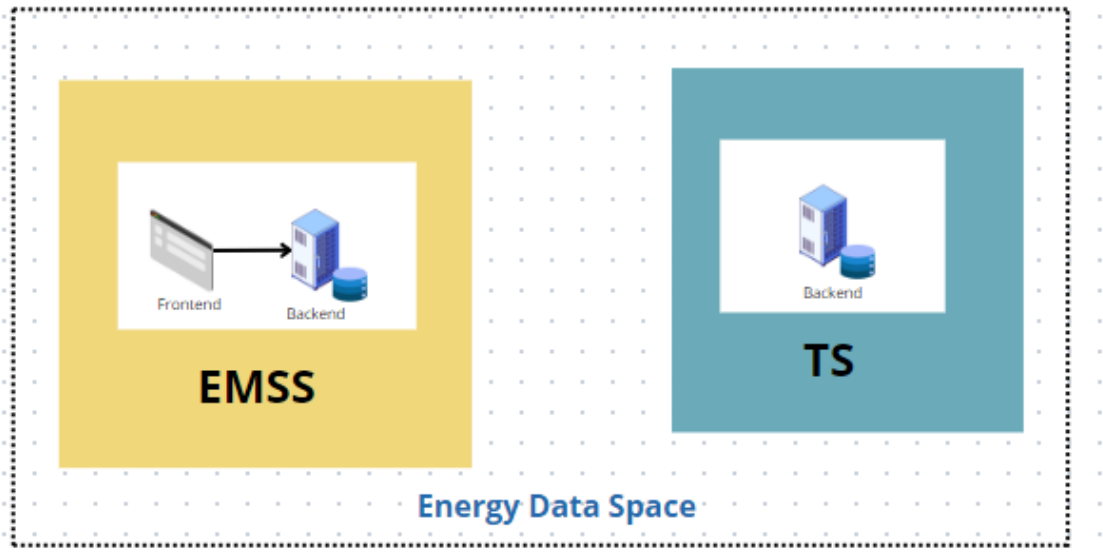


Figure 28 - Energy data space demo architecture

In the use case of energy data space, according to the document, the frontend of the system is realized using React, while its backend is by Express.js and MongoDB for the database. The choice of these technologies is driven by several factors, including being open-source and familiarity and experience with them.

Among the most influential and popular JavaScript libraries for user interface construction—especially single-page applications where dynamic and responsive user experience is principal in its requirement—is **React**. It achieves just that by providing an efficient way of handling complex UIs through its component-based architecture.

On the other hand, **Express.js** is a flexible **Node.js** web application framework that has a full set of features usable for web and mobile applications. Known for its simplicity and user-friendliness, it is an excellent option to implement APIs and handle HTTP requests.

Added to these is **MongoDB's NoSQL** database technology, serving flexible and scalable data storage. The most interesting thing about this schema-less database is handling the diversity of energies data structures, which are usually very different and therefore evolve into different data sources and formats.

To support this use case, the environment setup involves data from a Transform Monitoring Unit (TMU).

A TMU stands for a device or system that applies to monitor operational health and performance on a continuous basis from the power transformer. Transformers are those of an electrical power system, which is intended either for stepping up or stepping down the voltage levels for transmission or distribution. A TMU can gather information from multiple sensors. Our unit is composed by 2 equipment's the transformer itself defined as TRF G2 and the monitor DGA (Dissolved Gas Analysis). The data provided from a excel file although real data, we considered to be from a generic power plant.

After analyzing the structure of the excel file it was concluded that it has about 250.000 data entries that span around 9 months with the columns: source time, device, variable name, units, reason, type, acquisition mode and adjusted value.

- **Source time** – Is the timestamp that records the exact data and time of a particular measurement.
- **Device** – Is the device that was used in the measurement.
- **Variable name** – Indicate the parameter that is being measure.
- **Units** – Describes the units of measurement of the variable.
- **Reason** – Indicates the reason for the recorded data.
- **Acquisition Mode** – Describes how the data was acquired.
- **Adjusted Values** – Indicates the actual numerical reading of the measured parameter.

To make the data available in MongoDB, it was used a script in python that transforms the data into a JSON structure and inserts each operational event to the database.

The figure 29 illustrates the process of extracting the data from the excel file and import it into the mongo Atlas database.

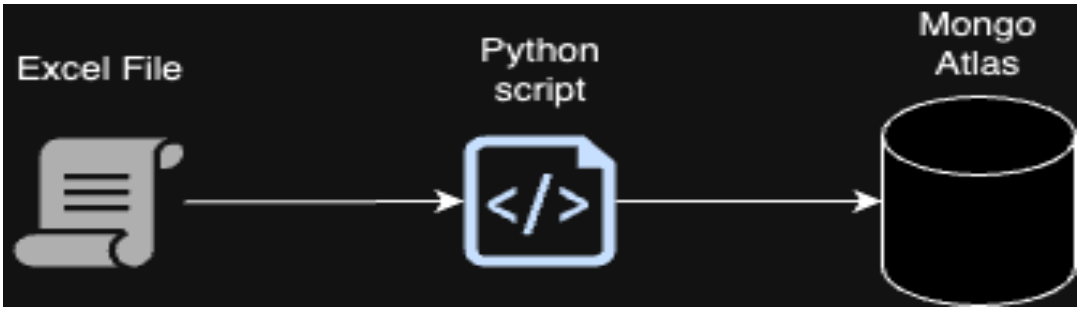


Figure 29 - Data extraction workflow

```python
from pymongo import MongoClient
import pandas as pd

df = pd.read_excel('Data - Operational Events.xlsx')

df.rename(columns={
    'Source Time': 'source_time',
    'Device': 'device',
    'Variable Name': 'variable_name',
    'Units': 'units',
    'Reason': 'reason',
    'Type': 'type',
    'Acquisition Mode': 'acquisition_mode',
    'Adjusted Values': 'adjusted_values'
}, inplace=True)

mongo_data = df.to_dict(orient='records')

client = MongoClient('mongodb+srv://admin:[redacted]@typo01.pb4kv2j.mongodb.net/?retryWrites=true&w=majority&appName=typo01')

db = client['transformer_data']
collection = db['operational_events']

collection.insert_many(mongo_data)

print("Data imported successfully!")
```

Figure 30 - Script to extract data from excel and add to a mongodb

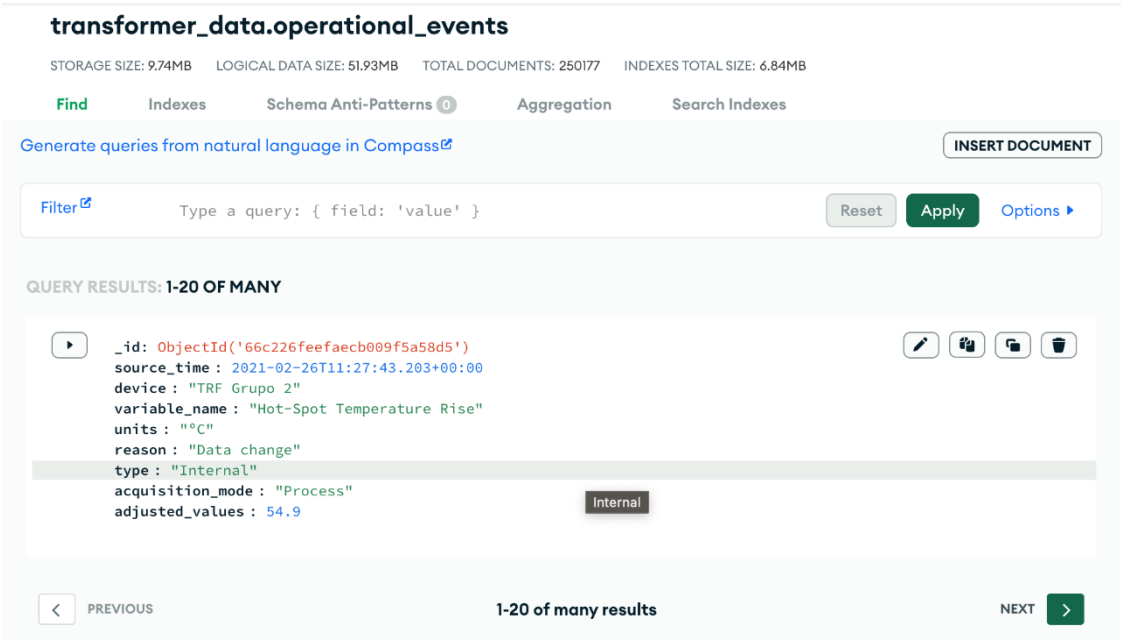


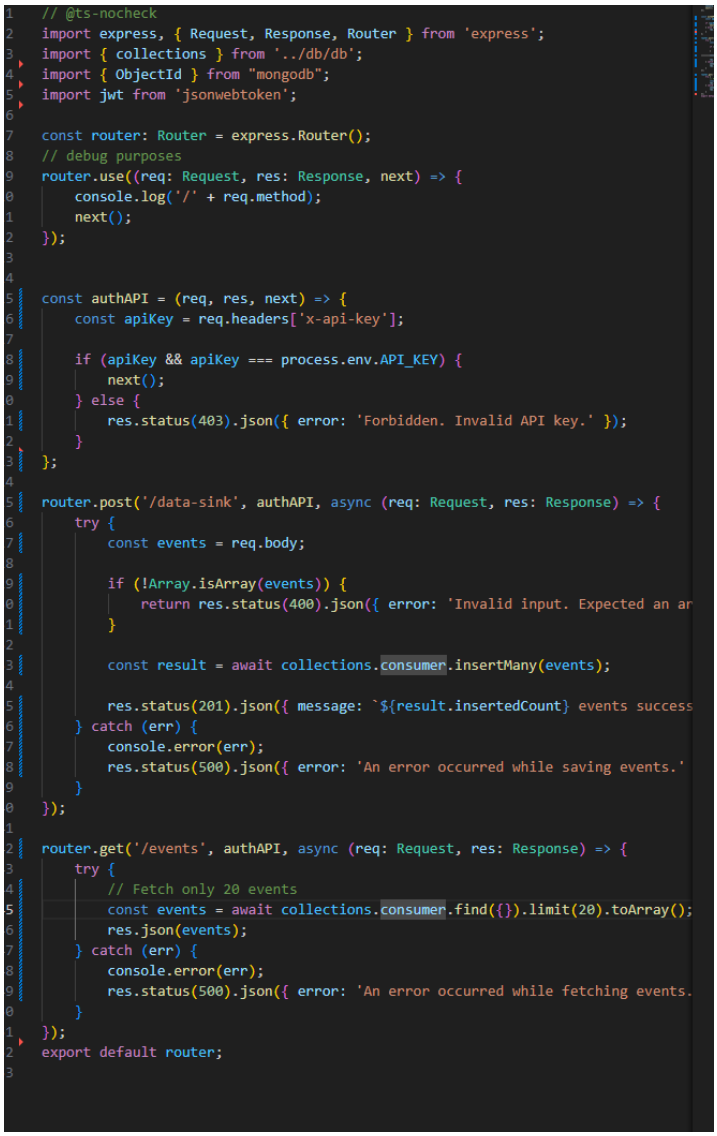
Figure 31 - Mongo Atlas

Once the data is secured in MongoDB, the next process is setting up a server hosting a REST API. This will give clients access to, and the possibility of interaction with, the data.

Figure 32 - Consumer REST API Server

However, since the Sovity EDC does not inherently support API access security features, it's crucial to implement additional security measures to protect the exposed API. In this scenario, we have the option to use either API key or API token (JWT) authentication to secure the API. For simplicity and ease of setup, we will opt for API key protection. API key authentication is straightforward to implement and provides an effective way to control access without the need for the more complex token management that comes with JWTs.

Now that the TS backend server is up and running, setting up a backend server for the EMS will be our next step, so it may serve the data in an efficient way. Besides, enhancing the UI as an improvement to the project will yield an easy and more intuitive way to interact with and visualize the data.



```
// @ts-nocheck
import express, { Request, Response, Router } from 'express';
import { collections } from '../db/db';
import { ObjectId } from "mongodb";
import jwt from 'jsonwebtoken';

const router: Router = express.Router();
// debug purposes
router.use((req: Request, res: Response, next) => {
    console.log('/' + req.method);
    next();
});


const authAPI = (req, res, next) => {
    const apiKey = req.headers['x-api-key'];

    if (apiKey && apiKey === process.env.API_KEY) {
        next();
    } else {
        res.status(403).json({ error: 'Forbidden. Invalid API key.' });
    }
};

router.post('/data-sink', authAPI, async (req: Request, res: Response) => {
    try {
        const events = req.body;

        if (!Array.isArray(events)) {
            return res.status(400).json({ error: 'Invalid input. Expected an ar
        }

        const result = await collections.consumer.insertMany(events);

        res.status(201).json({ message: `${result.insertedCount} events success
    } catch (err) {
        console.error(err);
        res.status(500).json({ error: 'An error occurred while saving events.'
    }
});

router.get('/events', authAPI, async (req: Request, res: Response) => {
    try {
        // Fetch only 20 events
        const events = await collections.consumer.find({}).limit(20).toArray();
        res.json(events);
    } catch (err) {
        console.error(err);
        res.status(500).json({ error: 'An error occurred while fetching events.
    }
});
export default router;
```

Figure 33 - Provider REST API Server

The EMS will act as a data sink for our scenario.

4.2.4 Use case workflow

In the section we will show how the process of consuming and offering data in a data space is made.

4.2.4.1 Offering data

The process contains the following three major steps: creating a data asset, creating a policy, and finally defining your data offer. All these steps are explained below.

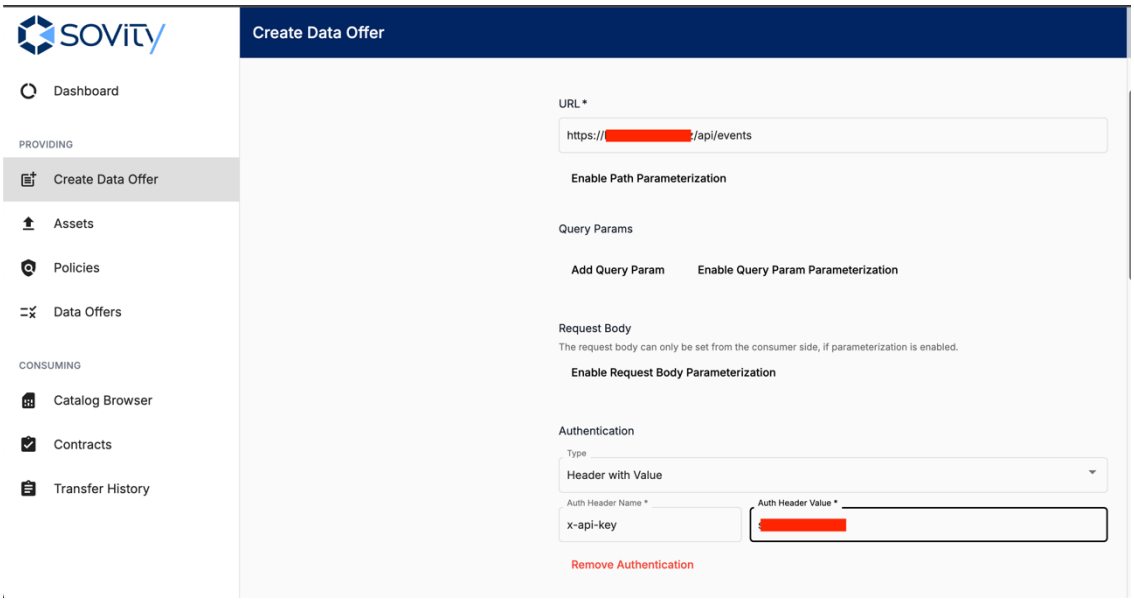


Figure 34 - Creating a data asset

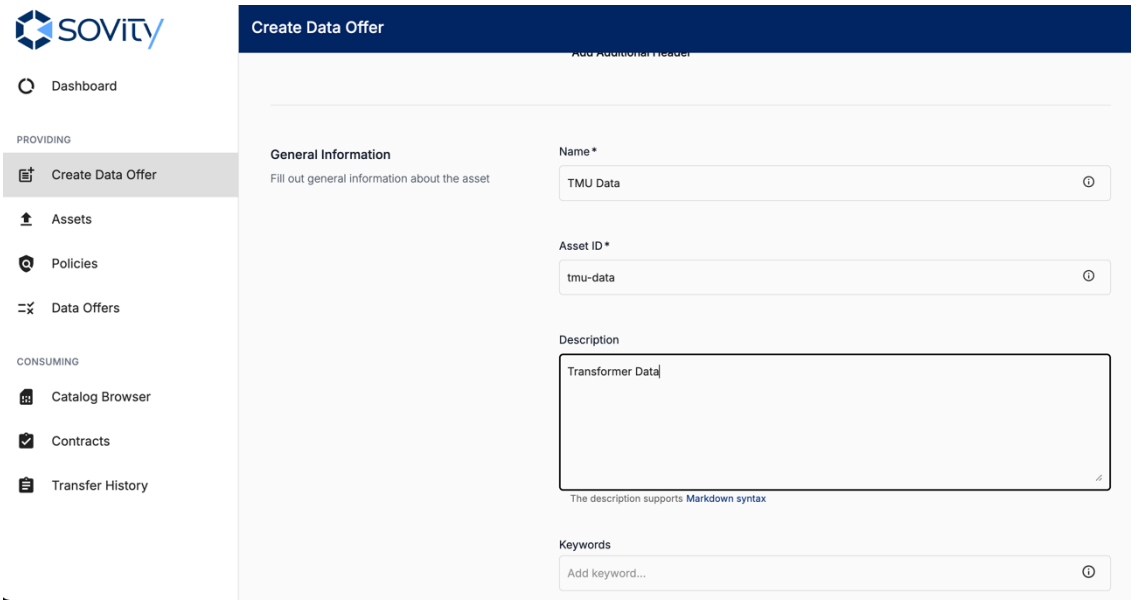


Figure 35 - Creating data asset continuation.

An employee utilizes the UI of their data connector to create a data asset. On the "Assets" page, the user inputs metadata about the data: name, version, description, keywords, etc. In addition, he provides certain technical information about this source.

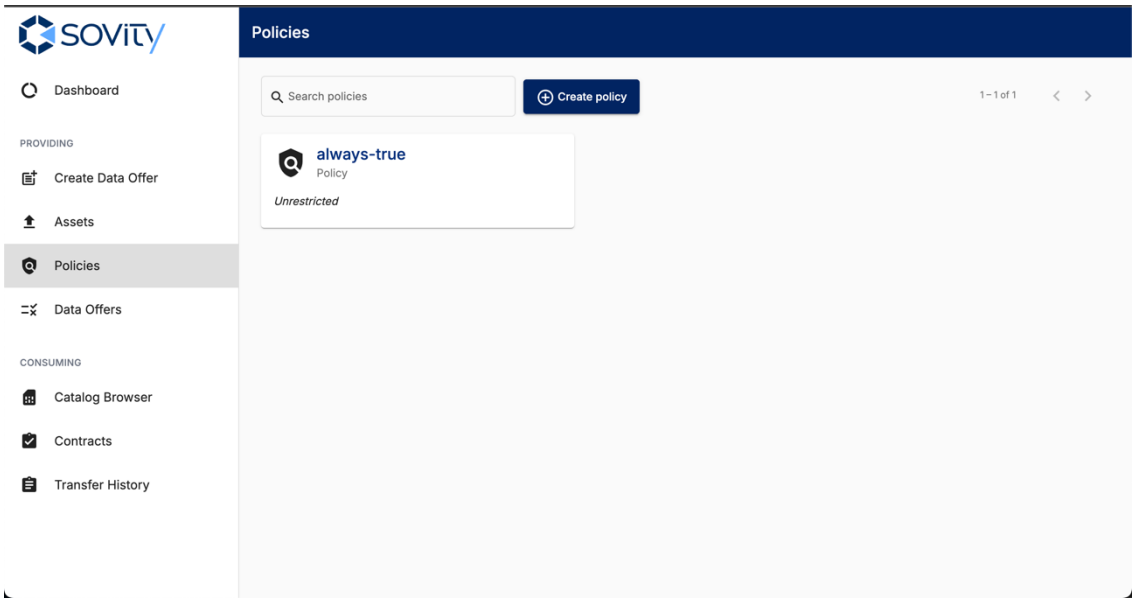


Figure 36 - Creating a policy

Three types of policies within the Sovity Community Edition EDC can be provided for the following:

- "Always-True": A general type of policy that would simply enable every participant to access the resource without any kind of restriction.
- "Connector-Restricted-Usage": The usage is restricted to specific connector instances.
- "Time-Period-Restricted": This constraint restricts use during a defined time.

An employee can create a new policy by visiting the "Policies" page and then clicking on "Create policy" using the connector UI.

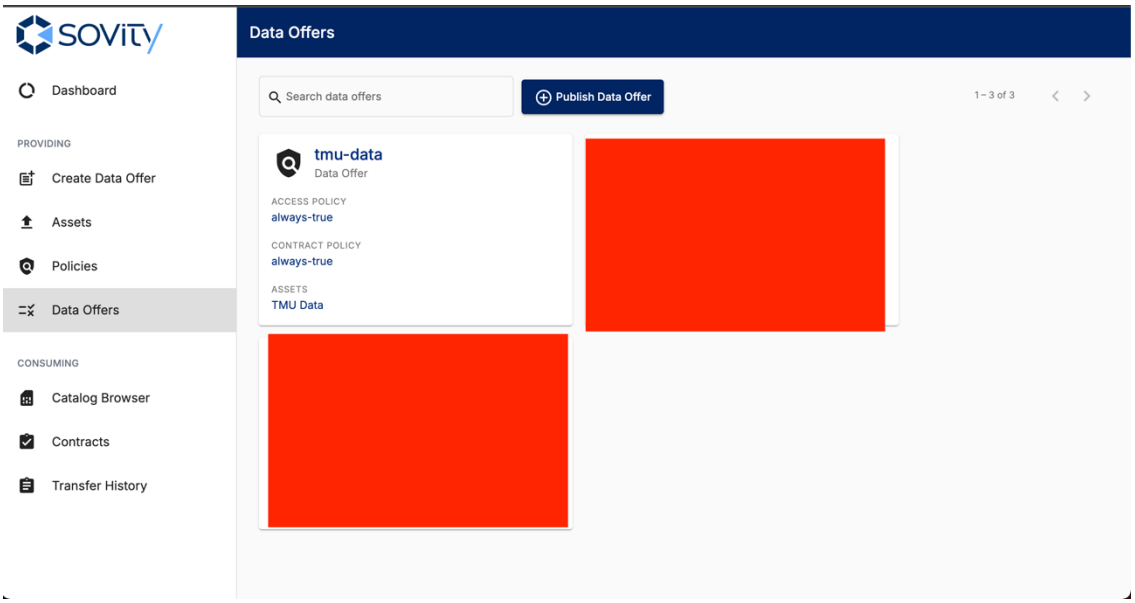


Figure 37 - Defining a data offer

In the case of contract definition, we must select the following two major policies:

- Access Policy: This is a policy describing how the offered contract definition shall be discoverable by the connectors of the other participants in the dataspace. It controls who can find the offer.

- Contract Policy: The policy that describes how the real data—in this case, the soil moisture measurements—can be consumed by the participants.

The employee creates a new Contract Definition via the UI of the connector and then goes to the page "Data Offers" and selects the option "Publish Data Offer". After the data offer has been created, this offer is published in the connector-based catalog of the provider. It makes this offer discoverable to other participants of the data space.

4.2.4.2 Consuming data

The process for data consumption involves three steps: data offer discovery, contract negotiation and establishment, and data transfer. In the next sections, we present each one of these steps in detail.

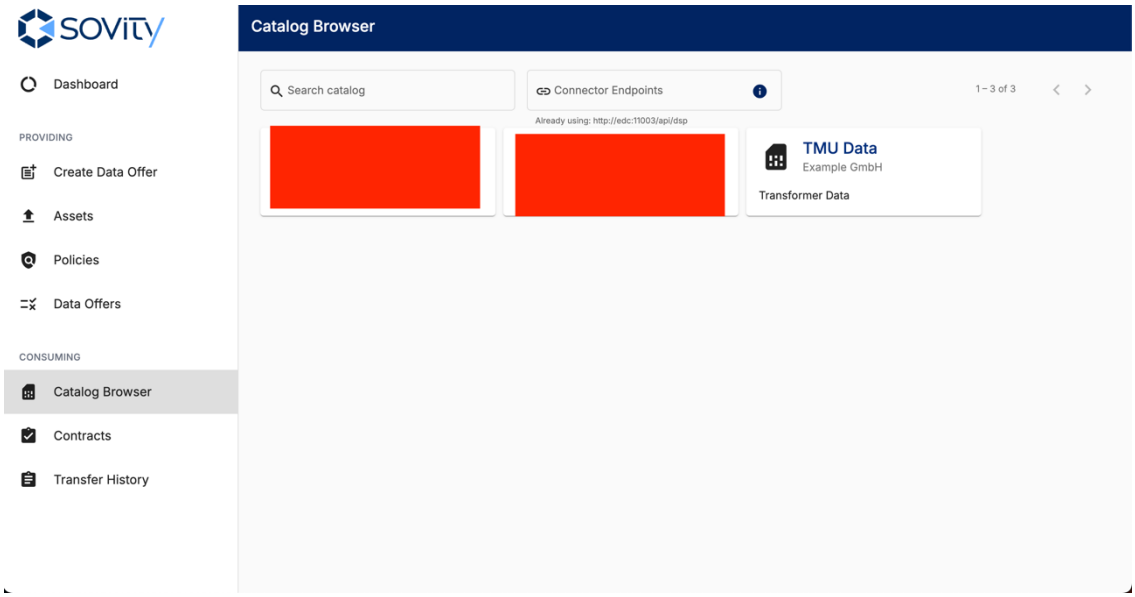


Figure 38 - Catalog Browser

To find corresponding data offers, a person will use their connector's UI to reach the page entitled "Catalog Browser". This now allows the employee to first judge whether the offer on the data meets his needs before further action in negotiation of the contract and transfer of data.

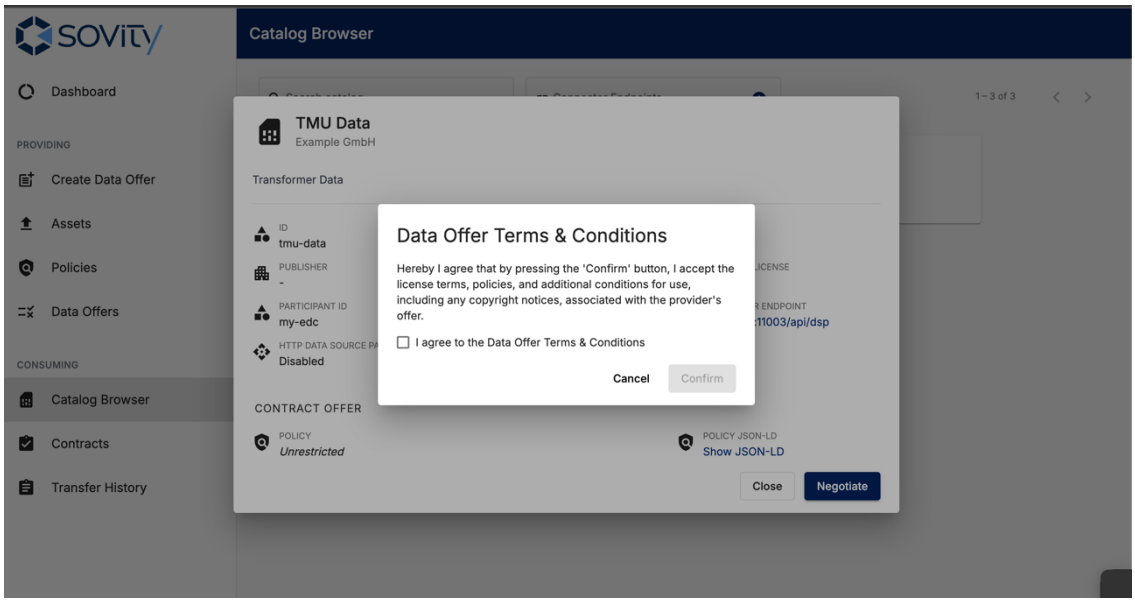


Figure 39 – Data offer negotiation

Having found the offer with the data, an employee of the provider can trigger the actual negotiation of the contract directly in their connector's UI from the detailed view of that offer, just upon the click of a dedicated button. After that, the data offer is negotiated between the connectors; there is no need for employees at either party to negotiate it manually. Once the negotiation is finally successful and both parties have agreed to the terms, a binding contract

shall be created and stored with both connectors. This clearly lays out the terms for sharing and use of the data.

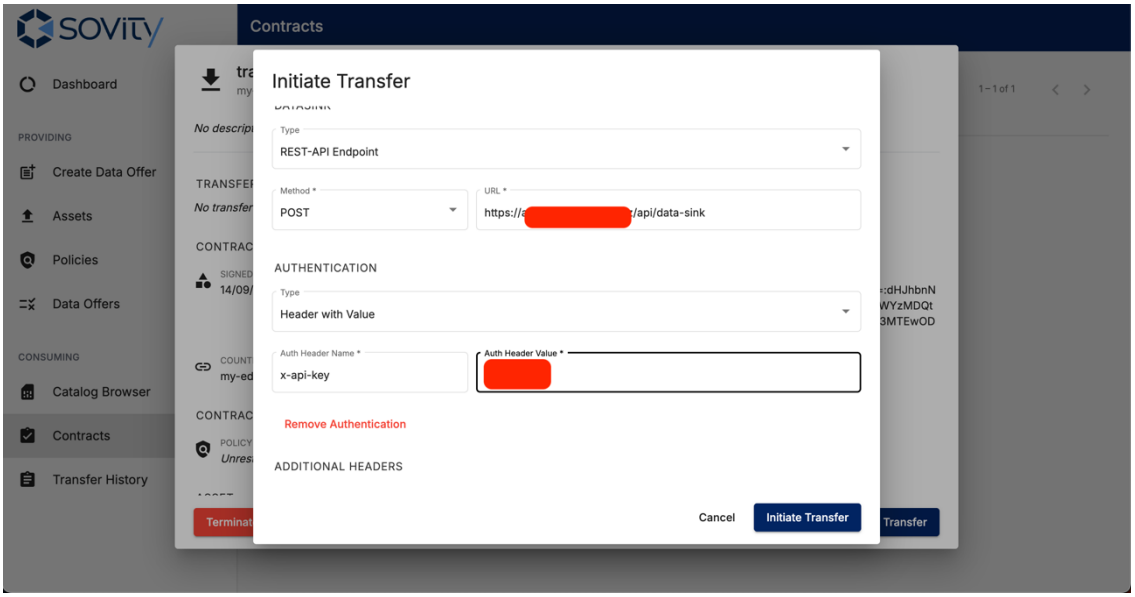


Figure 40 - Data Transfer

After a binding contract between participants has been established, an employee can trigger the actual data transfer for the operational events data directly in the UI of their connector. This button will, upon being clicked by the employee, prompt for details necessary to transfer the data to the specified HTTP endpoint address to which it is to be delivered and an API key header for authorizing the request. Upon initiation, the consumer connector sends a request for data transfer to the provider connector. This request refers to the contract (agreement) already established, therefore allowing the provider connector to verify if a contract is present and valid. Provided that the data transfer is granted, namely, a valid contract exists, and the attached policies are satisfied or can be satisfied, the provider connector proceeds with the retrieval of the data from the source previously established.

This will normally be realized by the provider connector sending an HTTP GET request to the provider backend to fetch the operational events data. Upon fetching, the provider connector then pushes the data into the consumer backend by sending an HTTP POST request to the specified backend endpoint of the consumer, thereby transferring data to the sink.

The image bellow show that the operation was successful, and the data was store in the consumer database.

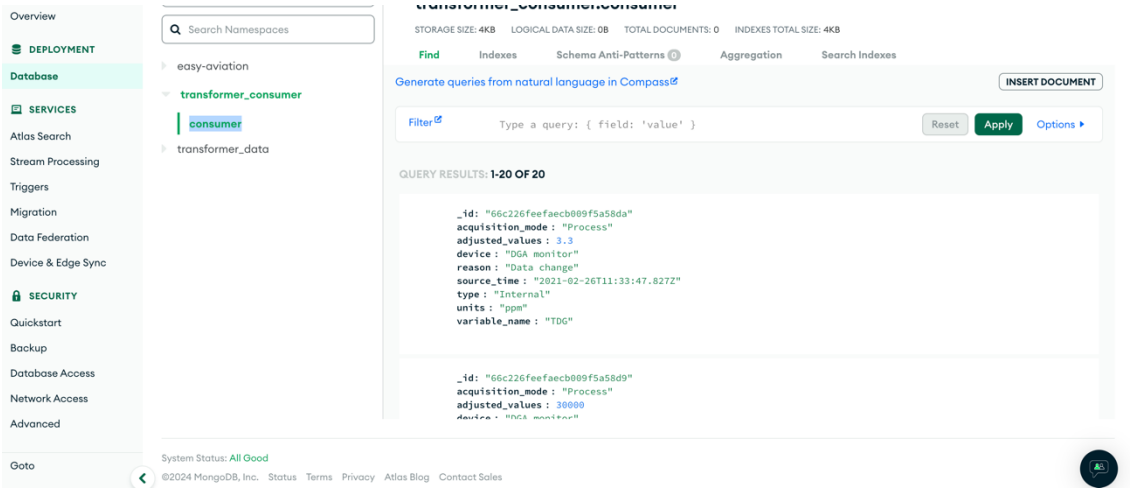


Figure 41 - Consumer Transformer Data

4.3 Critical Analysis

The primary objective of the Gaia-X framework is a transformational approach toward data sovereignty, security, and interoperability across all industries, decentralized infrastructure that will give a degree of control to stakeholders over their data while establishing a trusted digital environment with compliance.

Though the document we gain a sense of Gaia-X strength in areas such as data sovereignty and federation, the interoperability and compliance but it also rises few concerns.

A major challenge in Gaia-X is perhaps the complexity of its governance structure or model. The shared responsibility for maintaining compliance and trust among various participants can difficult the governance. When enforcing a standard to many sectors by means of ensuring accountability also can present a significant barrier.

When talking about interoperability, the integration of various cloud providers, for each different standards and technologies can be problematic specially when an address the topic of legacy systems.

One other aspect is for Gaia-X to succeed it needs a wider adoption where we can connect to its ability to scale. We already mention maintaining the structure is very complex and difficult and smaller businesses could have a hard time to integrate Gaia-X due to either entry costs or technical barriers.

Due to the limiting technical aspects of this task and the lack of extensive documentation, a demonstration of the functionality of Gaia-X in combination with the Sovity EDC was impossible within this thesis. Shifting the focus to create a minimum viable data space had to be considered, represented by a small-scale energy data space, to serve in the exploration of user interactions with it in reviewing user interaction with a generic architecture of a data space.

At the heart of the Gaia-X framework lies Gaia-X Clearing House or GXCH for short, which is supposed to handle every aspect related to trust and compliance issues within this network. However, poor documentation and hence technical support has led to crucial problems, which have been discovered in this thesis impeding the functionalities of the Clearing House. These challenge developers to a large degree when trying to ensure that a new implementation meets the underlined sovereignty and interoperability requirements of Gaia-X.

The issues that are experienced with the Gaia-X Clearing House show that while the architecture of this framework is good in theory, in practice it is undermined by a lack of operational clarity.

Finally, the use of Gaia-X in driving innovation within the energy sector will be analyzed, as well as its long-term viability as a digital infrastructure framework. In this regard, digital technologies and regulations are constantly evolving; for this reason, it is expected that the development of Gaia-X keeps up with these changes. Therefore, this section reviews the frameworks for integrating emerging technologies like AI, and their ability to ensure data sovereignty and interoperability principles over time.

5 Conclusion

This thesis has tried to explore the potential of the Gaia-X framework in view of development for secure, decentralized data infrastructures with a special view on application in the energy sector. While the analysis provided an extensive understanding of Gaia-X, architecture, main components, and working principles, it turned out in the practical phase of this work that the realization of a fully Gaia-X-compliant prototype system could not be realized within the given constraints.

The challenges also noted indicate a high level of complexity and emergence in the Gaia-X ecosystem. Despite this, the research still provides valuable deliberations on the foundational concepts of Gaia-X and what its implementation portends for data sovereignty, interoperability, and security across several industries. Although the realization of a functional prototype based on Gaia-X was not realizable, theoretical considerations taken up in the present thesis highlight even more strongly how important further development and cooperation within the Gaia-X community is to meet the challenges, accordingly, making the framework more suitable for practical work.

These results underline that Gaia-X tools and resources still require further development to make them usable and integrable into existing real-world systems, especially in business sectors like energy, where data management meets demanding requirements of security and efficiency. Future work should be directed at solving these technical and operatively relevant barriers detected in this thesis to enable Gaia-X to be successfully used in different industrial contexts.

The critical review underlines several strong points in the Gaia-X framework, such as high importance given to data sovereignty and interoperability, but also points out several challenges-for example, technical complexity and regulatory compliance issues-that must be resolved so that the full potential of Gaia-X will be reached across sectors. Meeting these challenges will be critical for enhancing its effectiveness and assuring scalability and usability for diverse applications.

Although Gaia-X revealed its little practical applicability in this study, the framework is still very promising for the future European data infrastructures. Further development and support of Gaia-X are indispensable if it should be developed into something useful for practical applications in the coming years.

5.1 Limitations

Although a Gaia-X prototype had set a desirable goal, it was not possible to give birth to such an artifact due to technical and operational limitations. There are strong links between these

limitations and the status of technologies and components involved in building a Gaia-X environment according to the last known specifications.

Gaia-X Digital Clearing House is, in general, in a very early stage of development. It lacks many basic functions not implemented or tested in production environments. This makes using it for building up a robust prototype quite a risk and unfeasible.

The documentation available on GXDC is very minimum and scant in crucial details, which creates problems while setting up, integrating, and testing the required components. It makes development slow and prone to errors because developers often must work on trial and error while facing some problem.

Three critical credentials that need to be obtained to validate the Gaia-X credential are as follows: participant, terms and conditions, and legal registration number. However, the creation and validation of the legal registration number for testing, like TaxID or EUID or EORI or VATID or LEI code, was quite a challenge. Without these validations, the credentials cannot consider being compliant with the requirements of Gaia-X; however, making a functional prototype is impossible without these.

Setting up the GXDCH for local testing presented several challenges, especially in terms of compliance with Gaia-X requirements. The likelihood of developing and testing a prototype was extremely low due to the lack of a suitable test environment and, above all, valid credentials.

Currently, none of the catalog services necessary to create a complete Gaia-X environment exist. Catalog services are essential for discovering and negotiating data offers within a Gaia-X dataspace. Without this, a huge part of functionality envisioned by the Gaia-X architecture could not be replicated.

6 References

Alethio Preukschat, Drummond Reed, 2021. Self-sovereign identity. Manning Publications.

Boris, O., Hompel, M., 2022. Designing Data Spaces. [WWW Document]. URL https://doi.org/10.1007/978-3-030-93975-5 (accessed 12.09.24).

Coelho, P., 2017. Internet das Coisas-Introdução Prática. Lisboa: FCA 158–178.

European Commission, 2024. EIDAS Regulation [WWW Document]. URL https://digital-strategy.ec.europa.eu/en/policies/eidas-regulation (accessed 9.12.24).

Gaia-X Association, 2021. The energy data space [WWW Document]. URL https://gaia-x.eu/wp-content/uploads/files/2021-06/Gaia-X_Data-Space-Energy_Position-Paper.pdf (accessed 12.09.24).

Gaia-X Association, 2024a. What is Gaia-X [WWW Document]. URL https://gaia-x.eu/what-is-gaia-x/ (accessed 01.09.24).

Gaia-X Association, 2024b. Gaia-X Architecture Document - 22.10 Release [WWW Document]. URL https://docs.gaia-x.eu/technical-committee/architecture-document/22.10/ (accessed 11.09.24).

Gaia-X Association, 2024c. Gaia-X Architecture Document - 22.10 Release [WWW Document]. URL https://docs.gaia-x.eu/technical-committee/architecture-document/22.10/ecosystem/ (accessed 14.09.24).

Gaia-X Association, 2024d. Gaia-X Digital Clearing House (GXDCH) [WWW Document]. URL https://gaia-x.eu/gxdch/ (accessed 08.09.24).

Gaia-X Association, 2024e. Gaia-X Compliance [WWW Document]. URL https://gitlab.com/gaia-x/lab/compliance/gx-compliance/-/blob/main/README-developer.md (accessed 08.09.24).

Gaia-X Association, 2024f. Gaia-X Wizard [WWW Document]. URL https://wizard.lab.gaia-x.eu/ (accessed 08.09.24).

Giussani Giulia, Steinbuss Sebastian, 2023. IDSA Data Connector Report WWW Document]. URL https://internationaldataspaces.org/data-connector-report/ (accessed 08.09.24).

International Data Space Association (IDSA), 2024. Data Spaces [WWW Document]. URL https://internationaldataspaces.org/why/data-spaces/ (accessed 02.09.24).

Maier, B., Pohlmann, N., n.d. Developing a Decentralised, User-Centric, and Secure Cloud Ecosystem Gaia-X secure and trustworthy ecoystems with Self Sovereign Identity [WWW Document]. URL [WWW Document]. URL https://internationaldataspaces.org/why/data-spaces/ (accessed 02.09.24).

Otto, B., Rubina, A., Eitel, A. et al., 2021. IDSA Position Paper: GAIA-X and IDS [WWW Document]. URL [WWW Document]. URL

https://www.researchgate.net/publication/348767747_GAIA-X_and_IDS (accessed 02.09.24.

Raghad Matar, Philipp Neuschwander, 2024. Realizing Agricultural Data Spaces with Eclipse Dataspace Components [WWW Document]. URL https://www.iese.fraunhofer.de/blog/realizing-agricultural-data-spaces/ (accessed 08.09.24).

Reiberg, A., Niebel, C., Kraemer, P., 2022. What is a Data Space? Definition of the concept Data Space [WWW Document]. URL https://gaia-x-hub.de/wp-content/uploads/2022/10/White_Paper_Definition_Dataspace_EN.pdf (accessed 08.09.24).

Š. Čučko, M. Turkanović, 2021. Decentralized and Self-Sovereign Identity: Systematic Mapping Study [WWW Document]. URL https://ieeexplore.ieee.org/document/9558805 (accessed 08.09.24).

Sovity, 2024. Sovity EDC Local Demo Deployment Guide [WWW Document]. URL https://github.com/sovity/edc-ce/tree/main/docs/deployment-guide/goals/local-demo (accessed 08.09.24).

Sovity GmbH, 2024. Sovity: Trustworthy Data Transfer with Data Spaces [WWW Document]. URL https://sovity.de/ (accessed 08.09.24).

Team Data Spaces, 2024. Joining Forces in 'Team Data Spaces' [WWW Document]. URL https://dataspaces4.eu/ (accessed 08.09.24).

Ulrich Ahle, H.B.K.B. et al., 2021. Design Principles for Data Spaces [WWW Document]. URL https://dataspaces4.eu/ (accessed 08.09.24).

Van Der Schaaf, H., Hertweck, P., Moßgraber, J., 2022. INSPIRE-Gaia-X Use-cases Final Report [WWW Document]. URL https://wikis.ec.europa.eu/download/attachments/65700040/INSPIRE-GaiaX-Use%20Cases%20Final%20Report_v1.1.pdf?version=1&modificationDate=1669628563800&api=v2 (accessed 08.09.24).

walt.id, 2022. Introduction to Self-Sovereign Identity [WWW Document]. URL https://walt.id/white-paper/self-sovereign-identity-ssi (accessed 08.09.24).

Wolford, B., 2024. What is GDPR? [WWW Document]. URL https://gdpr.eu/what-is-gdpr/ (accessed 9.12.24).

Zenza, F., 2024. Digital Twins no contexto da iniciativa Gaia-x dos data spaces. ISEP, Porto.