gaia-x Hub Austria

WHITEPAPER

Building a Dataspace: Technical Overview

Contact:

Veronika Siska *(veronika.siska@ait.ac.at)*

Vasileios Karagiannis *(vasileios.karagiannis@ait.ac.at)*

Mario Drobics *(mario.drobics@gaia-x.at)*

1 Introduction

1.1 Motivation

A large amount of data is generated in today's digitised world, but we cannot take full advantage of its value individually. Sharing data could fuel innovative data-driven applications, help to fulfil regulatory requirements, and generate monetary value. However, organisations hesitate to share their data for fear of losing control over who can access them and what they are used for. For data to be exchanged routinely among organisations, a collaborative environment—a so-called dataspace—is required in which participants can securely share and use data and related services, and where established technical and governmental standards ensure trust.

There are various initiatives developing software for dataspaces, but guidance on the capabilities, status, and usability of these available tools to implement dataspaces for specific purposes is lacking. The goal of this whitepaper is to bridge the gap between high-level concepts and framework-specific technical documentation and provide an overview of the current technical landscape of dataspace components. Since progress in many dataspace-relevant projects is rapid, this document aims to offer guidelines that help the reader become accustomed to the common terminology, the current state of the art, and the existing software applications.

1.2 Dataspace

A dataspace is a decentralised, open infrastructure for sovereign data exchange whose participants are aware and in control of the data they produce and consume as well as the involved services (hardware and software).

The concept of dataspaces appeared in computer science more than 15 years ago as a shift from central databases to storing data at the source[1]. Research at Fraunhofer ISST took this idea further, eventually leading to the establishment of the International Data Spaces Association[2] in 2015 and the creation of the initial concept and standards for dataspaces. Their reference architecture, the International Data Spaces Reference Architecture Model (IDS RAM)[3] describes the necessary components and their requirements in detail. Gaia-X[4] is a European initiative for the establishment of a federated and secure framework[5], [6] for sovereign data exchange. Gaia-X is taking the dataspace concept another step further by considering generic data-related services (e.g. storage, web servers) to enable interoperability between different cloud providers and IT infrastructures.

These different initiatives still have various technical and non-technical challenges to address. For example, while the software for exchanging data already exists, the processes and technical solutions to enable interoperability, ensure legal compliance, and establish trust in the ecosystem are still lacking. Business and governance models for the operation and use of such dataspaces are also still under development. Cooperation is the key to all areas, achieved through communities and initiatives spanning organisations in governments, academia, and the private sector. Many of the dataspaces currently developed aim to offer a federated, open infrastructure for sovereign data exchange based on shared standards and rules.

1.3 Gaia-X

The Gaia-X project aims to provide a framework in which all involved organisations from the various sectors can agree on a uniform ruleset, thereby ensuring that essential values are upheld and observed: data sovereignty, data privacy, confidentiality, security, technology neutrality, and interoperability. The goal of the Gaia-X project is to allow organisations, businesses, and users to process and exchange data efficiently and economically while still maintaining control over them—not only in terms of where those data are stored but also in terms of who may use them for which purposes. Gaia-X is therefore neither a new European data centre nor a new cloud service.

The Gaia-X project was initiated by 22 French and German companies and organisations in January 2021 and has since become a European initiative. Its organisation comprises three fundamental pillars: the Gaia-X European Association for Data and Cloud AISBL at the EU level, the national Gaia-X Hubs in multiple EU Member States and beyond, and the Gaia-X Community. Gaia-X was established as a not-for-profit association in Brussels under the name Gaia-X Association internationale sans but lucratif (AISBL); it currently has around 370 members (as of 03/2023).

In order to realise the principles and values of Gaia-X, appropriate specifications and software components referred to as the Gaia-X Framework[6] are currently being developed under the coordination of the Gaia-X Association. Their purpose is to allow organisations to exchange data and related services with one another while maintaining control over their use. In order to provide a fully transparent ecosystem, Gaia-X also aims to connect the data and infrastructure ecosystems—that is, the virtual data resources and underlying infrastructure services such as storage, computing or network solutions. Examples of specific services are the identification of participating organisations, the presentation of available data services, the automated monitoring of the observation of Gaia-X rules, or the notarial services for the preparation of contracts.

Gaia-X distinguishes between Federations and dataspaces. Gaia-X Federations are self-determined ecosystems of participants that can consume, produce, operate, or provide services, whereas a dataspace also includes specific services for data exchange such as connectors or usage policy enforcement. In this context, the goal of a Federation is to manage its participants and offerings, while dataspaces focus on sharing data in a specific area.

The Gaia-X Association publishes all specifications of the Gaia-X Framework[6], [7] and makes the software code for compliance available publicly under an open source licence: industrial enterprises, SMEs, start-ups, research facilities, public administration organisations, developers, and IT and cloud providers. All these stakeholders can present their products and services, exchange data, and jointly develop innovative business models based on the Gaia-X principles and with the help of the Gaia-X Framework.

The Gaia-X project is intended to promote innovation in all economic sectors and mitigate the increasing dependence on a handful of dominant organisations. While Europe may have already lost the race for the cloud infrastructure market, the cards for dealing in data are only just now being shuffled.

Gaia-X provides a framework embracing fundamental European principles: sovereignty, openness, fairness, security, and trust. It will therefore serve not just a scant few major players but instead enable countless companies and organisations worldwide to participate in the digital market more easily.

1.3.1 Gaia-X Framework

The Gaia-X Framework[6] builds on 3 conceptual pillars: Compliance, Federation, Services and Data Exchange. Gaia-X Compliance services are decentralised services that promote trust, while Federation services address interoperable and portable services for the ecosystem and data exchange services are responsible for the actual transactions including contracting, access and usage control, and logging. For each of these pillars, Gaia-X provides functional and technical specifications as well as open-source software components for compliance (see links on [6]). In addition, the Gaia-X Framework builds on the concept of the Gaia-X Digital Clearing Houses (GXDCH), which serve as execution nodes for compliance services.

1.4 International Data Spaces (IDS)

The first reference architecture approaches for sovereign data exchange were established by the International Data Spaces Association[2] founded in 2015. By now, the IDSA has more than 100 members, and its aim is to return control to data owners. In the IDS model, the data owner defines individual usage policies for its data assets, defining e.g. who can access the data and how they can be used. The International Data Spaces initiative is active in three main areas: research, standardisation, and providing software services and technology to the market.

The four main goals of IDSA are to build trust, guarantee security and data sovereignty, enable data ecosystems, and enable standardised interoperability. Every participant should trust the system as well as the other participants and be certain that data can only be shared under mutually agreed conditions and via trusted channels. In IDS, certification for each participant and software component ensures the trustworthiness of the system. Data sovereignty—meaning that participants are entirely self-determined with regard to their data—is the essential foundation and primary goal of the entire IDS architecture. Data sovereignty can enable new data ecosystems with their own cooperative business models. To enable such data ecosystems to interact with each other, interoperability is also a part of IDS. Therefore, the IDSA provides a comprehensive dataspace protocol[8] defining messages and API-bindings based on DCAT and ODRL.

1.4.1 IDS Reference Architecture Model

The IDS Reference Architecture Model[3] provides specifications and schemes for organisational roles and responsibilities as well as for technical components. It distinguishes four different business roles: Participants are users of the dataspace, involved in providing and consuming data; Intermediaries or "platforms" offer trusted services within the dataspace, such as storing and managing metadata or offering analytics services; Software Developers are IT companies providing software to participants of the dataspace; finally, Governance Bodies are responsible for creating trust and enhancing interoperability by providing, enforcing, and validating certifications and standards.

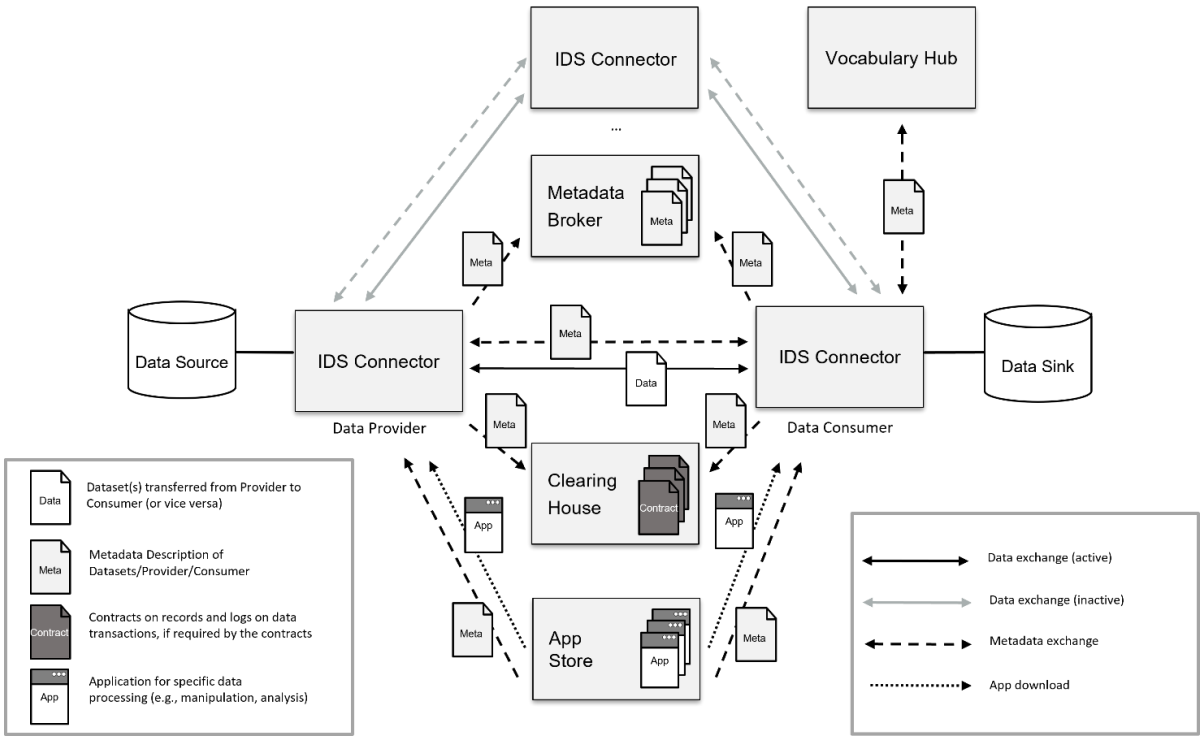

Figure 1: System diagram of the IDS RAM, from *Section 3.5: System Layer of the IDS RAM 4.0*.

On the technical side, the IDS architecture is based on the concept of Connectors for each participant that communicate with each other as well as with additional services common to the dataspace. Connectors act as secure gateways between entities and are responsible for exchanging data among each other. Processes within Connectors are separated into two areas: the control plane and the data plane. The control plane is responsible for all processes leading up to and following a transaction: identity and access management; handling offers; creating, negotiating, and settling contracts; logging. The data plane's sole responsibility is to transfer data after a successful contract negotiation (observed by the control plane). Components of the IDS RAM other than the Connectors include an Identity Provider for authentication and authorisation, a Metadata

Broker to interact with metadata in the catalogue, a Clearing House for logging transactions, and an App Store for additional data services.

Services in the IDS RAM are modular and can be combined and extended. For example, it is possible to take one of the IDS Connector implementations and connect it to a catalogue or an Identity Provider from a different stack. Given the comparatively long history of IDS, there are already multiple implementations for Connectors and other services, both under open source and proprietary licences, in use by all kinds of organisations from academia and NGOs to the industry.

The Dataspace Connector[9] was the first of the Connectors and served as the reference implementation for its successors. It was originally developed at Fraunhofer ISST and is now maintained by Sovity GmbH[10]. It has some limitations in terms of data exchange due to the coupling between the control and the data plane—that is, between processes before and after a transaction, like metadata exchange or contracting, and the actual transfer of data. However, it is a complete implementation with all components of the IDS RAM and was recently still in productive use e.g. by the Mobility Data Space[11] until April 2023. The Eclipse Dataspace Connector with its related Dataspace Components represents the successor of the original Dataspace Connector, with an architecture more suitable for production-level applications. It is discussed in detail in Section 3.1 of this whitepaper. For a detailed overview of available Connectors, readers are referred to the IDS Data Connector Report [12].

2 Dataspace architecture

There are various other initiatives related to dataspaces, but we will focus on tools for Gaia-X-compliant dataspaces, some of which are built using IDS-compliant tools. Various dataspace-related organisations provide specifications and standards to follow as well as reference architectures and minimal implementations. The latter are not intended 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, and we will discuss them in this section so as to have a common language to refer to when describing different dataspace components.

In general, a dataspace solution may consist of various elements (system components and participants), the most important of which are shown in Figure 2 and elaborated in the following subsections. They are:

- **Dataspace**: From a system perspective, a dataspace is the collection of necessary components that enable the sovereign exchange of data and services. As shown in the figure, multiple dataspaces can coexist under the same compliance service.
- **Asset Provider**: An individual or organisation offering an asset (e.g., a dataset, a service, etc.) that it holds/operates.
- **Asset Consumer**: An individual or organisation wishing to acquire/use an asset (e.g., a dataset, a service, etc.).
- **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 participants of the system.
- 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.

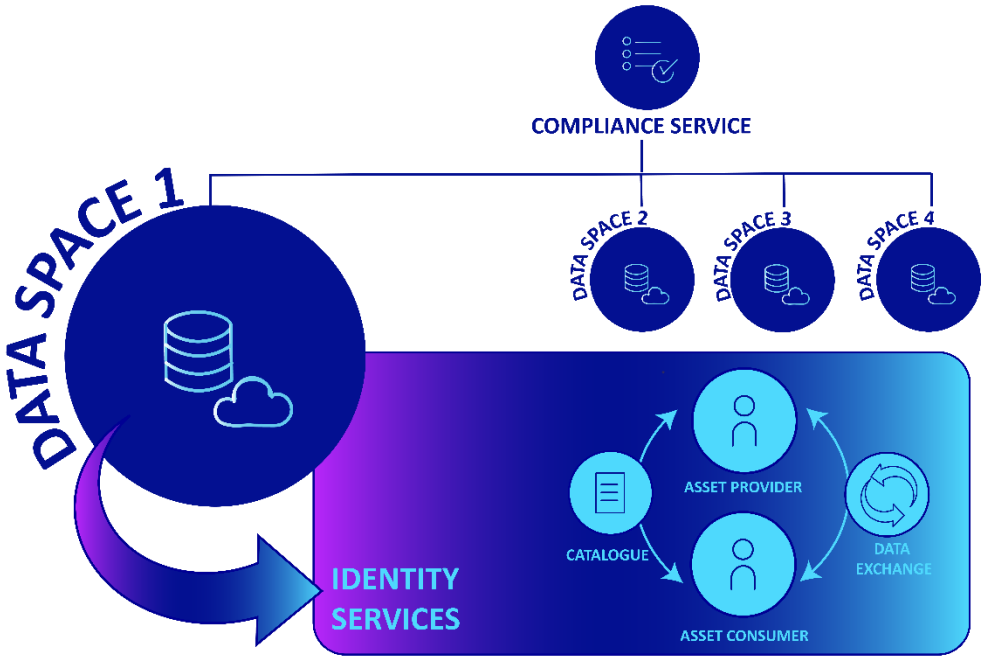

Figure 2 Basic elements of Gaia-X compatible data spaces (own visualization of AIT)

2.2 Data Exchange

Tools for the exchange of data or other services are an essential part of any dataspace. This can include **data transfer**, **contracting** between participants, defining and enforcing **usage policies**, **logging** transactions, **auditing**, and so on. Not all of these services are mandatory; requirements depend on the specific type of transaction. In the simplest case, a transaction may merely consist of agreeing on certain terms and then receiving a link to download a piece of data. On the other hand, a dataspace can also manage something as complex as rolling out services to a Kubernetes cluster and subsequently monitoring and managing them. The complexity and type of technical tooling should be adjusted to the needs of the dataspace and always use standardised, machine-readable descriptions to enable interoperability.

2.3 Catalogue

Likewise required is a **shared catalogue** providing standardised descriptions of offerings and allowing participants to search and select services. These descriptions, which are referred to as self-descriptions or metadata in the context of Gaia-X, should be extensible (so as to cater to the needs of specific industries) and verifiable (so as to help establish trust). Besides the shared catalogue of metadata, a dataspace can also offer services for **creating, signing, and managing standardised metadata**. Providers need to be able to offer services and control the visibility of their offers, and consumers need to be able to search and select services they are interested in. By using standardised metadata, such catalogues can even form a linked and/or nested network. The vision of Gaia-X is to collect all offerings in a shared, synchronised registry while also allowing additional aggregators at lower levels (centralised databases or decentralised storage e.g. on distributed ledgers or plain web storage) as well as visibility constraints to enable controlled or private dataspaces.

2.4 Identity

The next category of services is related to participants' identities: In order to securely identify them and specify their rights, they need to be **authenticated and authorised**. These services include the provision, management, and validation of identities based on an underlying trust framework as well as dataspace-

specific registries. To give participants control over their own identities and enable decentralised identities for services and assets, dataspaces often build on self-sovereign identities (SSI), although a federated approach or a central identity provider is also possible.

SSI can be implemented using Verifiable Credentials, as formulated in a W3C standard[13]; a form of digital certificates, containing cryptographically signed and thus tamperproof and automatically verifiable claims. In a dataspace, such credentials can be used to make trusted claims concerning the identity and attributes of participants, services, and offerings. It is essential that the participants agree on a set of issuers they trust and whose certificates they all accept. For SSI, additional components called **wallets** are required to manage (request, store, and present) credentials.

2.5 Compliance

To establish the basis of trust in dataspaces, a special set of services that **issue and verify certificates** are necessary to ensure compliance with regulations and standards. Compliance plays a role in the onboarding process of participants and services to an ecosystem and must remain verifiable at all times. This trust layer forms the basis for all dataspaces based on a given framework (e.g., Gaia-X), but dataspaces can also add additional rules and checks.

In the Gaia-X Trust Framework[7], certificates that signal compliance can be awarded to any entity (participant, resource, service offering) in the form of verifiable credentials. These certificates of compliance are compulsory in order to be part of the Gaia-X ecosystem, and additional labels certifying compliance with specific rules (e.g., European Control, Art. 6 GDPR) are attainable for service offerings that have been audited and vetted. Gaia-X offers two main components for compliance: the Gaia-X Registry holding compliant services and participants, and the Compliance API, where certificates can be obtained and verified. Additionally, at the dataspace level, software components as well as operational environments can be certified, e.g. for the IDS standards using the IDS Certification Scheme [14], [15].

It is important to note that services for compliance with a given framework (e.g. Gaia-X) need to be provided by the respective framework (but not necessarily operated by it), while dataspaces are required to use such services to be compliant themselves. Specifically for Gaia-X, in the latest Gaia-X Trust Framework, Gaia-X Digital Clearing Houses (GXDCH) will serve as the places to obtain compliance, with independent instances operated separately. The dataspace components that enable compliance by connecting to the compliance services (particularly to those of Gaia-X) will thus be part of our technology stack–specific discussion, but not the compliance services themselves.

2.6 Additional components

A dataspace can include further components to facilitate usage and provide domain-specific additional functionalities. A **portal** is essential for user-friendly access to the dataspace, including access to the participants' accounts and the provision and consumption of services. **Orchestration** of the ordered services or **data processing** applications can also supplement the core functionalities; they can be tailored to the needs of the participants and the types of exchanges in the system. Other plugins can be used to **enforce rules**, such as usage policies or legal regulations.

3 Dataspace stacks

In the following section, we will present the three major stacks that can be used build a dataspace: the Eclipse Dataspace Components, the Gaia-X Federation Services, and the Gaia-X Web3 Ecosystem. For an overview, see

Table 1. We will focus on the available dataspace components in the respective stacks and their underlying technology as well as on practical implications (advantages and limitations).

	EDC	*GXFS*	*Pontus-X*
Technology	Modular components built with Java & Gradle	Modular components built primarily with TypeScript	Blockchain-based technology built on Ocean Marketplace
Website	https://projects.eclipse.org/projects/technology.edc	https://www.gxfs.eu	https://portal.minimal-gaia-x.eu
Repository	https://github.com/eclipse-edc	https://gitlab.com/gaia-x/data-infrastructure-federation-services	https://github.com/deltaDAO/mvg-portal
Sample projects	Catena-X, Eona-X, Health-X dataLOFT, etc.	Planned adoption by German Gaia-X supported projects	EuProGigant, moveID, Berlin State Library, etc.

Table 1: Basic facts for the dataspace stacks described in this whitepaper: the Eclipse Dataspace Components (EDC), the Gaia-X Federation Services (GXS), and the Pontus-X Gaia-X Web3 Ecosystem & GEN-X Network (Pontus-X). Note: Project lists are not exhaustive.

3.1 Eclipse Dataspace Components

The Eclipse Dataspace Connector[16] and the related dataspace components offer a modular, extensible framework based on the IDS framework while also offering Gaia-X compatibility. In contrast to its predecessor, the Dataspace Connector, the control and data planes in the Eclipse Dataspace Connector are separated and thus offer better scalability. Processing data transfer separately is better suited to handling large datasets or complex protocols as well as reusing already available technologies within organisational infrastructures.

The project is hosted by the Eclipse Foundation and used by a number of dataspaces such as the Gaia-X lighthouse projects Catena-X[17] and Eona-X[18]. It is compatible with the IDS RAM but also actively developed to be compliant with Gaia-X standards, particularly the Gaia-X Trust Framework[19]. EDC components are written in Java and built with Gradle, but they can be embedded in any form of application. The EDC project also includes a separate repository for setting up a minimal viable dataspace (MVD) with EDC components in order to support technical adoption and onboarding. An overview of the architecture is shown in Figure 3.

The Connector can handle identity and access management (IAM) modules in three flavours: for the OAuth protocol, for decentralised identities (DID), and for the Dynamic Attribute Provisioning Service (DAPS) by IDS. To manage identities for a single entity (e.g., serving certificates in the case of DID), an identity hub is also available as a separate component to be run in the Connector environment. EDC also offers catalogue services as part of the Connector, which can store records of its own data assets as well as crawling and caching external ones. Furthermore, there is a separate service for managing dataspace membership, called the Registration Service, as well as a Trust Framework Adoption layer to allow configuration of the Gaia-X Trust Framework and corresponding policies.

The data plane is implemented as modules of the Connector, but can (and in a productive environment, should) be run in a separate environment. There are extensions for different protocols, data sources, and transfer mechanisms. This currently includes file transfer via HTTP as well as connections to the three major blob storage services of cloud providers: AWS, Azure, and Google Cloud. As with all EDC services, there is the option of writing custom extensions tailored to the needs of specific dataspaces.

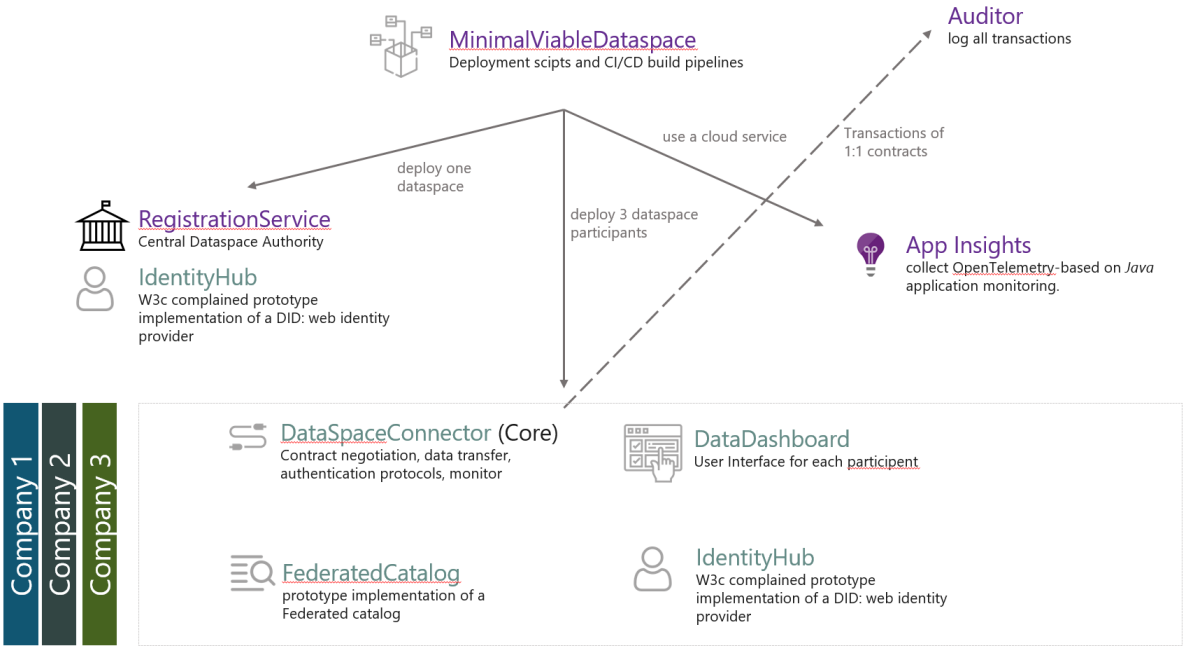

Figure 3: Architecture of the Eclipse Dataspace Components, using the example of the Minimal Viable Dataspace. Graphic provided by the EDC maintainers.

3.2 Gaia-X Federation Services

The Gaia-X Federation Services (GXFS)[20] represent the minimum technical requirements and services needed to operate federated Gaia-X ecosystems of infrastructure and data. GXFS is funded by the German Ministry for Economic Affairs and Climate Action (BMWK) based on a decision by the German Parliament. The project was established to kickstart the development of Gaia-X Federation Services as specified in a joint community process. The eco – Association of the Internet Industry with its head office in Cologne, Germany, acts as procurer to coordinate the specification work and award contractors for the implementation based on an open EU-wide tender process. The repositories are currently maintained by the Gaia-X Association but transfer to the Eclipse Foundation is planned for 2023.

GXFS are based on the Gaia-X architecture currently at version 21.03., but adaptations to the new Trust Framework in version 22.04. are in progress. GXFS offers specifications as well as baseline open-source code for the minimum set of services necessary to operate Gaia-X Federation Services (see Figure 4). Such ecosystems consist of interconnected data and infrastructure ecosystems aggregated in so-called Federations that are individually orchestrated and operated with the help of Federation Services. It serves as a reference implementation, but other implementations are possible. GXFS only includes the services necessary to operate a federation, with a generic scope: They consider not only data exchange, but also other types of offerings such as cloud services, physical machines, or algorithms. Executing data exchange, for example via connectors, is outside the scope, however. The services of GXFS form a modular toolbox from which Federations can pick and use what they need, adapting elements as necessary.

To ensure transparent and trustworthy transactions, GXFS offers Data Sovereignty Services – a Data Contract Service to create and manage contracts and a Data Exchange Logging Service to track transactions and enable audit trails for contracting parties. The Federated Catalogue is a searchable, up-to-date catalogue of metadata

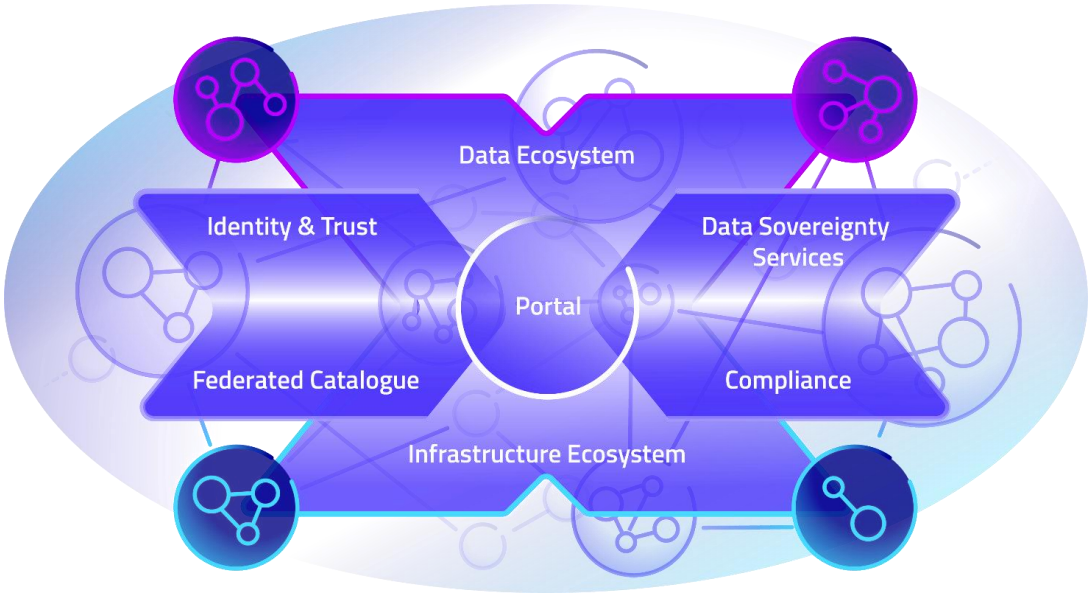

Figure 4: GXFS infrastructure components from the working groups Identity & Trust, Federated Catalogue, Portal, Data Sovereignty Services, and Compliance connecting the infrastructure and data ecosystems. Graphic provided by eco – Association of the Internet Industry.

(self-descriptions) shared within a Federation, and GXFS also offers additional tools for creating, visualising, and validating such self-descriptions for participants and assets.

Identities in GXFS are handled by the Authentication & Authorisation Services, which also include the management of decentralised identities and validation of participants during the onboarding process. Separate credential managers for organisations and natural persons are also provided, along with trust services to ensure trust for participants and services using cryptographic proofs. Compliance services are related to these identity services: a Continuous Automated Monitoring Service to monitor compliance of self-descriptions in the federated catalogue and a Notarization API to create verifiable representations based on asset and participant self-descriptions.

GXFS also provides additional services for supporting participants: a portal with a UI to integrate all services and an Orchestration Service to instantiate and manage infrastructure services such as virtual machines from the Federated Catalogue search results via the portal.

3.3 Pontus-X

A slightly different approach to the technical implementation of a Gaia-X ecosystem is taken in Pontus-X, the Gaia-X Web3 ecosystem and GEN-X network built by deltaDAO AG. This ecosystem is rooted in Web3 technologies, using a blockchain and smart contracts as a secure, distributed storage. The components build on the Ocean Protocol and the Polygon Labs software stack to provide the federation services required by dataspaces[21] (for the architecture, see Figure 5). The stack is used by used by Gaia-X lighthouse projects, i.e., EuProGigant[22] and moveID[23], and other European initiatives. All components are provided as open-source software under an Apache 2.0 license.

The current implementation uses a custom Ethereum Virtual Machine (EVM)–compatible blockchain, the GEN-X network[24], built with the Polygon Edge framework to be transitioned to Polygon Supernets. To add resilience to the network, it is distributed across European cloud service providers as well as geographically;

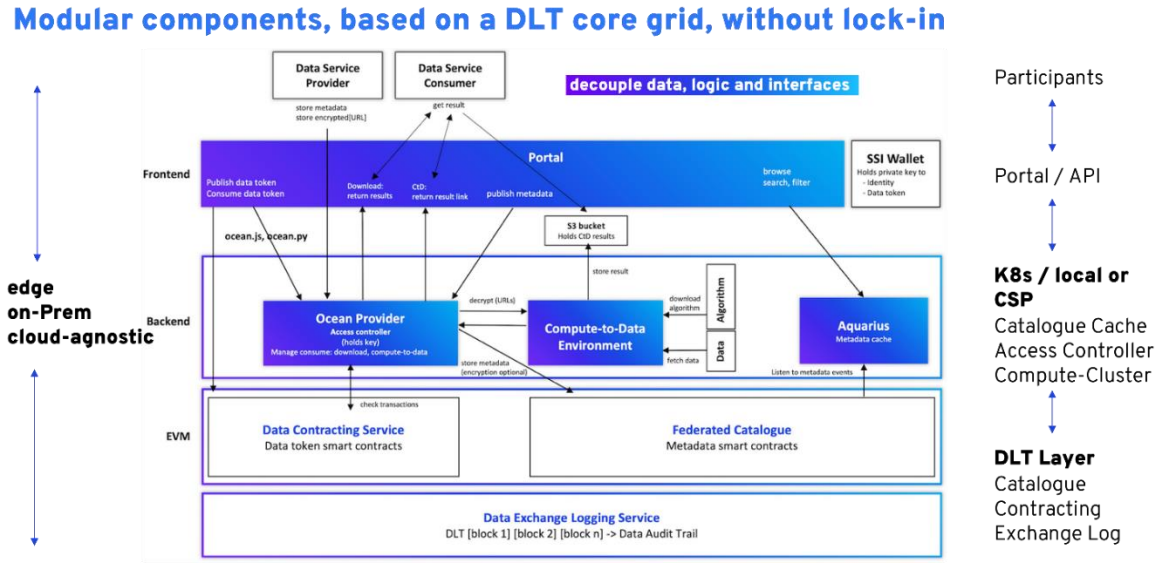

Figure 5: Example architecture for a service provider in the Gaia-X Web3 ecosystem. Graphic provided by deltaDAO AG.

meanwhile, any of the underlying services can be run by individual service providers. The GEN-X chain uses Proof of Authority (PoA), an eco-friendly and scalable consensus mechanism where approved accounts ("validators") are allowed to verify transactions and create blocks. The network is public-permissioned, which means that community-approved and identified federators can join. There are currently eleven validators, all based in Europe: Arsys (Spain), BigchainDB (Germany), deltaDAO AG (Germany), EuProGigant (Austria/Germany), Exoscale (Switzerland/Austria/Germany), Ionos (Germany), Staatsbibliothek zu Berlin (Germany), Wobcom (Germany), Software AG (Germany), and TU Wien (Austria) and Universitat de Lleida (Spain), with more to come.

The shared blockchain also acts as an interoperability layer between dataspaces and participants in the Gaia-X Web3 ecosystem: They can be seen as subsets of participants and offerings with their own portals, infrastructures, attached networks, and optional additional rules. The catalogue can be stored on-chain or off-chain in the form of (optionally encrypted) metadata for each service offering, which are accessed by the portal via a cache (Aquarius) or directly from the network. The catalogue can either store the full metadata and Gaia-X compliant self-descriptions in individual ERC721 smart contracts, or as pointers to a webspace or decentralised off-chain storage.

Contracts defining the terms of use for assets are recorded on-chain as smart contracts and as part of the service offering's self-descriptions, which can again be stored off-chain or on-chain. Smart contracts are automatically executed if consumer and service provider sign an agreement and when the conditions are met. Usage restrictions (e.g., download or use for computation only, contract expiry, licensing, access controls, and pricing/settlement) can be part of these contracts. Audit trails are enabled by logging all transactions on the blockchain.

Services for SSI-based authentication and authorisation are also included, such as Role-Based Access Controls (RBAC) and fine-grained permissions. A fully-fledged SSI integration including the processing of REGO policies through SSI verifiers is currently under development. Furthermore, the Pontus-X ecosystem requires all participants to comply with the Gaia-X Trust Framework by providing Gaia-X compliant self-descriptions verified via the Gaia-X compliance service.

The underlying framework, Ocean Protocol, supports static and dynamic data access as well as running computations on data without directly accessing them, instead only receiving the results ("compute-to-data"). The latter paves the way for the preservation and enhancement of data privacy, on-demand aggregation/anonymisation along with federated analytics and learning.

Monetisation and instant settlement are an integral part of the stack: Access to a given offering is managed through its own specific ERC20 utility token ("data token"). These data tokens can currently be acquired in exchange for other ERC20 tokens (i.e., OCEAN tokens, EURO tokens), or in the future through off-chain payments e.g., via credit card or SEPA.

3.4 Summary

Any of the outlined stacks can be used to establish a dataspace, including by way of a mix of these ready components from the above-described stacks and custom additional software. A summary of the available components for each stack is shown in Table 2.

	EDC	GXFS	Web3
Identity & Trust	IAM modules (oath, did, daps)	Authentication & Authorisation Services	DID on GEN-X network
	Identity Hub	Personal Credential Manager	SSI wallet
		Organisational Credential Manager	
Catalogue	Connector catalog	Federated Catalogue	Federated Catalogue: metadata smart contracts on-chain
	Federated catalog	Self-description wizard	Aquarius: cache for metadata
Exchange	Data plane with extensions for different protocols	Data Contract Services	Data contracting Service: data token smart contracts
	Control plane for contracts	Data Exchange Logging Service	Data Exchange Logging Service: audit trail on-chain
			Compute-to-data environment
Compliance	Trust Framework Adoption	Validation as part of authentication/authorisation	Validation as part of an SSI layer
		Continuous Automated Monitoring Service	
		Notarisation Service	
Other	Data Dashboard (only for demonstration)	Portal	Portal
	Registration Service	Orchestration Service	

Table 2: Main components for each technology stack. Only dataspace-specific services are listed under "Compliance", not those provided by the framework (e.g., Gaia-X).

4 Interoperability

Currently, interoperability between different ecosystems/stacks is limited, particularly at the level of software components. However, common standards for the description of participants and services ensure a level of compatibility: All ecosystems understand the same machine-readable metadata. A further pillar of interoperability lies in the common trust framework: Services can be certified by the same authority and share compliance services. Trust anchors for identification can also be shared so that participants can be identified in different dataspaces using the same credentials.

Interoperability is an important goal of all organisations participating in the development of distributed data economies. European organisations are striving to define ways of complementing business and technical convergence as well as the interaction between different architectures (e.g. Gaia-X and IDS[25]). The Data Spaces Business Alliance (DSBA) formed by the Big Data Value Association (BDVA), FIWARE Foundation, Gaia-X, and the International Data Spaces Association (IDSA) drives the adoption of dataspaces across Europe and recently initiated the development of a common framework for dataspaces, the Minimum Viable Framework (MVF)[26]. iSHARE, a trust framework for dataspaces, recently also began a collaboration with the goal of

aligning the trust frameworks of iSHARE and Gaia-X. Meanwhile, the Data Spaces Support Centre was funded by the European Commission as part of the Digital Europe Program with the aim of facilitating common dataspaces and providing a dataspace blueprint.

5 Conclusion

Openness and collaboration are key for successful dataspaces: The individual technical implementation needs to meet the needs of the participants and conform to agreed rules and domain-specific legal or business requirements. It is also essential to contribute back to the community and release software developed in dataspace projects as open source. Joining the communities of the corresponding open-source projects is beneficial not only for dataspace developers but also for users, to receive support and shape future development. This facilitates the existing active, open community as well as technical convergence for a truly open data economy.

The status of available dataspace software is changing rapidly to meet the needs of the increasingly mature initiatives. We can already see a few dominant technologies emerging and converging towards shared standards stemming from the best practices seen in early dataspace projects. The next technical steps for widespread adoption will bring easy-to-handle software packages for those aiming to host their own dataspace and managed services for out-of-the-box projects. Acquiring proof of compliance and (optionally) additional labels for each offering will increase trust in the system, and a shared, distributed catalogue of offerings will enable openness and interoperability between different ecosystems.

6 Glossary

Dataspace

The term “dataspace” refers to a type of data relationship between trusted partners adhering to the same high standards and guidelines regarding data storage and exchange. However, a critical aspect of the dataspace notion is that data are not stored centrally but instead at their source and are thus only transferred as necessary.

A dataspace is the sum of all its participants – which may be data providers, users, and intermediaries. Dataspaces can be nested and overlapping, so that a data provider, for instance, can participate in several dataspaces at once. Data sovereignty and trust are essential for dataspaces to work and support relationships between participants. Each dataspace provides specific data and thereby forms a solid ground for each ecosystem. The software required to implement dataspaces runs on cloud/edge cloud infrastructures.

When organisations establish a Gaia-X Federation, an associated dataspace is created. Gaia-X-compatible dataspaces facilitate cooperation between different actors, thereby generating innovation opportunities and enabling value creation.

Gaia-X Federation

The term “Federation” refers to a set of multiple actors jointly adhering to the Gaia-X rules and using the same standards and defined interfaces to ensure secure and sovereign exchange of data. A Federation thus also describes transactions between different IT systems. It reduces uneconomical island solutions and monopolisation while supporting interoperability and the building of trust.

Organisations may agree to establish any number of Federations defined according to Gaia-X standards. They do so on the basis of mutually agreed rules and with the help of the technical Federation Services provided by Gaia-X. This allows them to connect their IT services in a trustworthy manner to mutually exchange data and use them securely.

IDS Connector

The term “IDS Connector” refers to a service acting as a secure gateway between entities. Connectors are responsible for exchanging data among each other as well as for communicating with other dataspace services. A Connector can be considered a data exchange service and includes its own identity, which can be registered to the dataspace on behalf of its owner. A Connector is typically deployed at a participant’s IT infrastructure and communicates with that participant’s internal data storage services. It is responsible for exposing the participant’s assets to the dataspace via its internal catalogue as well as for usage policies that can be associated with those assets. The Connector also handling contract negotiation on behalf of the participant and orchestrates processes to execute data transfer after the successful establishment of a contract.

Verifiable Credential

The term “verifiable credential” (VC) refers to digital certificates holding claims together with cryptographic proof of their correctness, making them tamperproof and automatically verifiable. The W3C open standard for verifiable credentials[13] defines the data model and format for VCs. Such a VC consists of the credential metadata (e.g., issuance data), the credential subject (the data to be verified), and the proof (cryptographic signature and its metadata). It uses the JSON linked data (JSON-LD) syntax so that context data can be included, for example about the entity that the credential refers to.

Verifiable credentials enable self-sovereign identities, where a user is in full control of their identity; this makes them particularly suitable for dataspaces. The verifiable credential flow describing the process for managing and using VCs consists of two parts: one for obtaining and one for using credentials. To obtain a new VC, the owner submits their claims to a trusted issuer, who is registered with their public key in a publicly available, verifiable data registry. The issuer signs the claim and sends it back to the owner. The owner stores their credentials, possibly from different issuers, in their personal wallet. When a service requires the owner to provide proof of a certain claim, the owner can prepare a verifiable representation of their available credentials, potentially combining and selectively disclosing them. The service can then verify this representation by looking at the proof(s) and validating that they were indeed signed by the trusted issuer by looking at their record in the verifiable registry of issuers.

In the case of dataspaces, the participants can decide which issuers they accept. The Gaia-X trust services also act as an issuer themselves, for example issuing Gaia-X membership credentials for participants and compliance credentials for services. Gaia-X self-descriptions are also verifiable credentials with attributes described by the Gaia-X Trust Framework[7].

7 Acknowledgements

We thank Markus Spiekermann (Fraunhofer ISST) for his contribution, particularly regarding IDSA and EDC, including the graphic for EDC; Kai Meinke (deltaDAO AG), Thomas Komenda (deltaDAO AG), Meike Molitor (deltaDAO AG), Frederic Schwill (deltaDAO AG) and Albert Peci (deltaDAO AG) for their input especially for Pontus-X, including the corresponding architecture figure and Vivien Witt (eco - Verband der Internetwirtschaft) for her contribution particularly regarding GXFS, including its graphic. Without their input for the discussed technological stacks, this whitepaper would not have been possible. We are also thankful for Michael Fälbl (Association Industry 4.0 Austria) and Georg Hahn (OSSBIG Austria) for their review and Athina Lyoku (AIT Austrian Institute of Technology) for designing the graphic showing the basic elements of dataspaces (Figure 2).

References

[1] M. Franklin, A. Halevy, and D. Maier, 'From databases to dataspaces: a new abstraction for information management', *SIGMOD Rec.*, vol. 34, no. 4, pp. 27–33, Dec. 2005, doi: 10.1145/1107499.1107502.

[2] 'Home - IDSA', *International Data Spaces*. https://internationaldataspaces.org/ (accessed Jan. 02, 2023).

[3] B. Otto, S. Steinbuss, A. Teuscher, and S. Lohmann, 'IDS Reference Architecture Model', Zenodo, Apr. 2019. Accessed: Jan. 02, 2023. [Online]. Available: https://zenodo.org/record/5105529

[4] 'Home - Gaia-X: A Federated Secure Data Infrastructure'. https://gaia-x.eu/ (accessed Jan. 02, 2023).

[5] 'Publications - Gaia-X: A Federated Secure Data Infrastructure'. https://gaia-x.eu/mediatech/publications/ (accessed Jan. 06, 2023).

[6] 'Gaia-X Framework'. https://docs.gaia-x.eu/framework/ (accessed Jan. 25, 2023).

[7] 'Gaia-X Trust Framework - 22.04 Release'. https://docs.gaia-x.eu/policy-rules-committee/trust-framework/22.04/ (accessed Jan. 26, 2023).

[8] 'Introduction - Dataspace Protocol'. https://docs.internationaldataspaces.org/dataspace-protocol/overview/readme (accessed Mar. 06, 2023).

[9] 'Home', *Dataspace Connector*. https://international-data-spaces-association.github.io/DataspaceConnector/ (accessed Jan. 05, 2023).

[10] 'sovity - The IDS Operating Company', *sovity*. https://sovity.de/ (accessed Jan. 05, 2023).

[11] 'Mobility Data Space – the Data Sharing Community'. https://mobility-dataspace.eu/ (accessed Jan. 05, 2023).

[12] International Data Spaces Association, 'Data Connector Report'.

[13] 'Verifiable Credentials Data Model v1.1'. https://www.w3.org/TR/vc-data-model/ (accessed Jan. 05, 2023).

[14] N. Menz, A. Resetko, and Prof. Dr. B. Otto, 'Framework for the IDS Certification Scheme 2.0', Zenodo, Jan. 2019. doi: 10.5281/zenodo.5244858.

[15] S. Steinbuss, N. Menz, A. Resetko, and J. Winkel, 'IDS Certification explained', Zenodo, Nov. 2019. doi: 10.5281/zenodo.5675945.

[16] M. T. Delgado, 'Eclipse Dataspace Components', *projects.eclipse.org*, Jun. 16, 2021. https://projects.eclipse.org/projects/technology.edc (accessed Jan. 05, 2023).

[17] 'Home | Catena-X'. https://catena-x.net/en/ (accessed Jan. 05, 2023).

[18] 'Data-sharing on mobility in Europe', *EONA-X*. https://eona-x.eu/ (accessed Jan. 05, 2023).

[19] 'Trust Framework Adoption'. Eclipse Dataspace Components, Jan. 16, 2023. Accessed: Mar. 06, 2023. [Online]. Available: https://github.com/eclipse-edc/TrustFrameworkAdoption

[20] 'Gaia-X Federation Services', *GXFS.eu*. https://www.gxfs.eu/ (accessed Jan. 02, 2023).

[21] 'Ocean Protocol Use Case: Gaia-X, A Federated European Data Infrastructure'. deltaDAO, Sep. 22, 2022. Accessed: Jan. 25, 2023. [Online]. Available: https://github.com/deltaDAO/Ocean-Protocol-Use-Cases/blob/3517eeeadc8bbf58626c4611d631179de9c883a6/Ocean%20Protocol%20Use%20Case%20-%20Gaia-X%20v2.pdf

[22] 'EuProGigant - EuProGigant'. https://euprogigant.com/en/ (accessed Mar. 10, 2023).

[23] 'moveID', *moveID*. https://moveid.org/ (accessed Mar. 10, 2023).

[24] 'Introduction | GEN-X Network'. https://docs.genx.minimal-gaia-x.eu/docs/intro/ (accessed Jan. 26, 2023).

[25] Prof. Dr. B. Otto, 'GAIA-X and IDS', Zenodo, Jan. 2021. doi: 10.5281/zenodo.5675897.

[26] Data Spaces Business Alliance, 'Technical Convergence Discussion Document'. [Online]. Available: https://data-spaces-business-alliance.eu/dsba-releases-technical-convergence-discussion-document/

About the Gaia-X Hub Austria

Accelerating European Data Ecosystems
for Economic, Ecological and Societal Value Creation

The Austrian Gaia-X Hub was launched in March 2022 on the initiative of the Federal Ministry of Finance (BMF), State Secretariat for Digitalisation, and the Federal Ministry for Climate Protection, Environment, Energy, Mobility, Innovation and Technology (BMK) (www.gaia-x.at). As a national Gaia-X focal point for companies, public institutions and initiatives, the hub aims to ensure the implementation of the Gaia-X strategy in Austria and thus guarantee data sovereignty as a prerequisite for sustainable economic growth and social justice.

To support the Gaia-X vision effectively, visibly, and sustainably, the Gaia-X Hub Austria pursues the following four strategic areas of activity for all involved or interested Austrian organizations and individuals:

No. 1: Provide information to establish trust

The Gaia-X Hub Austria will establish itself as the first point of contact for Gaia-X in Austria. It will ensure all work results, guidelines, documentation, and open-source codes are made accessible to all interested stakeholders in Austria in a timely, uncomplicated, and comprehensible fashion.

No.2: Facilitate entry to support development

The Gaia-X Hub Austria will support Austrian organizations in their adoption of data-based business models and enable Austrian businesses of all sizes to initiate or participate in innovative implementation projects. This will allow them to obtain experience, acquire know-how, and increase their competitiveness on the European market.

No. 3: Concentrate forces to increase effectiveness

The Gaia-X Hub Austria will help organizations to network efficiently and establish independent work groups and concrete alliances for implementation. The Hub will facilitate access to national and international funding programmes and showcase successful Austrian implementation projects in other EU countries. The Gaia-X Hub Austria will also ensure the greatest possible synergy with relevant Austrian initiatives and programmes.

No.4: International networking to enhance visibility

The Gaia-X Hub Austria will establish a strong network with other national Hubs – especially with those in neighbouring countries and Eastern European states. The purpose of this powerful concerted voice is to ensure that topics and questions particularly relevant to Austria are heard at the European level, that Austrian projects are appropriately presented, and that the Austrian influence on decisions pertaining to the data economy at the EU level is increased.

Gaia-X
Innovation through
digital sovereignty

Gaia-X European
Association for Data
and Cloud AISBL

info@gaia-x.at
www.gaia-x.at

Place of publication:
Vienna, March 2023

AIT Austrian Institute of
Technology GmbH

Giefinggasse 4
1210 Wien
Austria