piveau-X: A Compliance-Focused Semantic Web-based Catalog for Data Spaces

Fabian Kirstein[1], Michael Gysel[1]

[1]*Fraunhofer FOKUS, Berlin, Germany*

Abstract

Data Spaces are a rapidly evolving concept that enable sovereign and secure data sharing. A Data Space consists of multiple building blocks and follows a well-defined governance and security framework. One of these key building blocks is a catalog for data discovery, which allows participants to publish and find data, services, and other artifacts in a compliant and trustworthy manner. Additionally, many Data Space standards, such as Gaia-X and Verifiable Credentials, utilize Semantic Web technologies as a data modeling language. With piveau-X, we introduce a catalog solution for Data Spaces that elevates the Semantic Web, enabling the seamless application of formal Data Space specifications. Our solution employs a Triplestore as the primary database and SHACL as the data modeling language, ensuring maximum flexibility in managing data and metadata structures while also facilitating the dynamic reuse and integration of existing ontologies. Furthermore, piveau-X natively integrates with identity and attestation services, making it a comprehensive solution for creating Data Space catalogs. Its built-in SPARQL and Elasticsearch-powered search and query capabilities allow for precise browsing of data artifacts via APIs or the web frontend. We successfully validated piveau-X's practical applicability in the Gaia-X research project POSSIBLE through five real-world use cases.

Keywords

Data Space, Gaia-X, RDF, SHACL, Catalog

This demo paper integrates with the W3C Dataspaces Community Group by addressing the generic dataspaces issue https://github.com/w3c-cg/dataspaces/issues/2.

1. Introduction

Data Spaces represent a thriving and widely adopted concept that facilitates sovereign and secure data and service sharing. Unlike centralized data integration mechanisms, Data Spaces keep data at its source without transforming it into a unified format. Integration occurs semantically via a shared vocabulary and metadata model. Furthermore, the concept introduces a business and legal layer that ensures trust, interoperability, portability, and data sovereignty. [1]

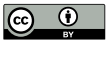
A variety of technical and organizational building blocks are established to instantiate Data Spaces. One of the central building blocks is a catalog that enables the publication and discovery of data, services, offerings, and other artifacts. Another key building block is a trust framework, which ensures the integrity and authenticity of offered resources by leveraging trust anchors, compliance rules, and digital identities. [2]

We propose that implementing the core concept of Data Spaces and their central building blocks can be achieved through the full adoption of Semantic Web technologies, such as RDF, SHACL, and SPARQL. While these technologies are already applied in various flavors and applications of Data Spaces, they are not yet utilized to their full potential. In this demo paper, we introduce our Semantic Web-based catalog solution, piveau-X, which aims to maximize the use of Semantic Web technologies while ensuring compliance and trust. Our solution offers the following core features: (1) the use of a Triplestore as the primary database, enabling high flexibility and rich query capabilities via SPARQL, (2) a flexible and dynamic metadata model configuration based on SHACL, combined with full-text search powered by Elasticsearch, (3) a user-friendly web frontend for filtering and searching assets, and

The Third International Workshop on Semantics in Dataspaces, co-located with the Extended Semantic Web Conference, June 01, 2025, Portorož, Slovenia

fabian.kirstein@fokus.fraunhofer.de (F. Kirstein); michael.gysel@fokus.fraunhofer.de (M. Gysel)

0000-0002-9064-2546 (F. Kirstein); 0009-0001-6821-8685 (M. Gysel)



(4) direct integration with identity and attestation services. Our solution is built on the open-source catalog software piveau [3]. We successfully demonstrated its practical feasibility within the scope of a Gaia-X research project, covering five real-world use cases.

2. Background and Related Work

This section provides an overview of the foundational concepts that underpin our research, including, Gaia-X, Decentralized Identifiers (DIDs) and Verifiable Credentials (VCs), and Data Spaces Support Centre (DSSC).

Gaia-X is a European initiative for an interoperable, federated data infrastructure that ensures data sovereignty for its participants. By establishing a framework built on European principles of data security, GDPR compliance, transparency, and interoperability, Gaia-X enables organizations to share data securely and under rules that preserve control for data owners [4]. Within this framework, standardized resource types—namely, Gaia-X Legal Participant, Service Offering, and Data Resource—are defined to consistently represent organizational identities, digital service offerings, and data assets in a compliant and interoperable manner [5].

As part of this framework, DIDs are used to establish digital identities for organizations and VCs are used for authentication and authorization purposes. DIDs are W3C standardized self-sovereign identifiers that allow entities to autonomously create, manage, and update cryptographically verifiable digital identities without centralized authorities [6]. Moreover, W3C-standardized VCs provide tamper-evident, cryptographically secured credentials from trusted authorities, enabling privacy-preserving, machine-verifiable claims of identity or qualifications [7].

The DSSC is an initiative funded by the European Commission that aims to identify and define standards, best practices, technologies, and tools to build and foster European Data Spaces. The DSSC supports public sector entities and businesses in creating and managing these Data Spaces, thereby accelerating the digital transformation across various domains. [8]

Piveau, a comprehensive and scalable data management catalog, was extended to meet the requirements of a Data Space catalog. The solution is based on Semantic Web technologies, utilizes a Triplestore as its primary database, and Elasticsearch as its indexing and search service. Furthermore, piveau is primarily designed to support Open Data platforms. [3]

In addition to these foundational technologies, several related works have explored practical implementations of Gaia-X compliant catalogs. The Eclipse Cross Federation Services Components (XFSC) Catalog was developed as a reference implementation of the Gaia-X federated cloud service architecture. It serves as a directory of services and providers, managing metadata records, schemas, and VCs while enabling service searchability, comparability, and data sharing. To support domain-specific requirements, it allows the submission of custom metadata schemas defined using OWL, SHACL, or SKOS. Furthermore, it performs its own syntactic, cryptographic, and schema verification checks on VCs, as opposed to using the Gaia-X Digital Clearing House (GXDCH). Notably, the XFSC Catalog avoids storing RDF data altogether, instead relying on openCypher's labeled property graph (LPG) data model for storing and searching data. [9]

Aster-X, created under the French Gaia-X Federation Services initiative, also makes use of the Eclipse XFSC. Developed as a Gaia-X Data Space demonstrator, the Aster-X catalog showcases functionalities such as Self Sovereign Identity (SSI) wallets, SSI-based permissions, automated contract negotiation, data exchange, and direct integration with the GXDCH for compliance verification. Similar to the XFSC Catalog, Aster-X stores data using openCypher's LPG data model. [10]

Unlike XFSC-based Gaia-X catalogs, Prometheus-X implements its own open-source building blocks to facilitate data sharing with the Gaia-X ecosystem. More specifically, Prometheus-X comprises a set of core services to facilitate participant onboarding, service discovery, contract negotiation, consent management, and direct integration with the GXDCH for compliance verification. [11] Furthermore, Prometheus-X circumvents RDF support by storing data and services in JSON-LD in a MongoDB database, and provides reference data models to users [12].

Pontus-X is the Gaia X web3 ecosystem that uses Distributed Ledger Technology (DLT) to store records of data sharing contracts, settlements, and usage rights, and enforces them via smart contracts. Pontus-X makes use of the Ocean Protocol to store metadata on-chain, connects the metadata to a database, and integrates the GXDCH for compliance verification. Furthermore, the Pontus-X catalog focuses on publishing a wide variety of software services, and thus does not support extensible data schemas. [13]

3. Requirements and Architecture

The primary objective of piveau-X is to provide an out-of-the-box solution for the Data Space building block **Publication & Discovery** [2]. However, the design of piveau-X comprises of more facets. Our solution includes broader aspects based on both conceptual and technical concepts of Data Spaces and real-world implementations of data infrastructures. Furthermore, our solution is built around Semantic Web technologies, as they provide a well-established and mature foundation for developing interoperable and harmonized data management solutions. The conceptual aspects are based on the Data Spaces Blueprint v1.5 [2]. Based on the blueprint, we define the following requirements: (1) cover interoperability, sovereignty, and value creation; (2) reuse established standards; and (3) demonstrate integration into an existing Data Space.

In the following sections, we present the detailed design decisions and overall architecture of piveau-X. Figure 1 illustrates the architecture, depicting all components and services. From a high-level perspective, we divide our architecture into two realms: the Data Space Realm and the piveau-X Realm.

The Data Space Realm provides essential Data Space services required for a functional catalog. A Decentralized Identity Service manages the digital identities of participants and authorizes their interactions within the Data Space. An Attestation Service verifies cataloged resources against compliance rules and digitally signs them. Data Models define the format, structure, and constraints of resource descriptions. Finally, an independent Verification Service ensures the integrity of published resources.

The piveau-X Realm, which consists of the catalog, is divided into three layers: Configuration, Databases and Integration, and Main Services.

The **Configuration layer** establishes a connection to the Data Space Realm. SHACL files serve as a direct representation of the Data Models, defining the supported data structures of the catalog. The Identity Keys are cryptographic private keys linked to the Identity Service and, consequently, to a specific participant. The Profile represents the general runtime configuration of the catalog.

The **Databases and Integrations layer** comprises two main databases: a Triplestore for managing and storing RDF and an Elasticsearch search server for enabling high-performance full-text search. This layer also includes a Signing Service, which is responsible for cryptographically signing resources using the provided private keys. This functionality is offered as an independent service to enable self-sovereign key management and signing, allowing on-premises deployment when using a central catalog instance.

The **Main Services layer** consists of the core services. The Search Service provides an interface to the Elasticsearch server, the Repository functions as middleware for the Virtuoso Triplestore, and the Web Frontend delivers a user-friendly interface.

Table 1
Mapping of Data Space Pillars to piveau-X Design

Data Interoperability	Data Sovereignty and Trust	Data Value Creation Enablers
SHACL-based data validation	Pluggable signing service	Full-text search engine
Support for external data models	Integration with attestation services	SPARQL query interface
Triplestore as primary database	DID service support	Web-based frontend

Table 1 illustrates how the various design aspects of piveau-X align with the technical pillars of the

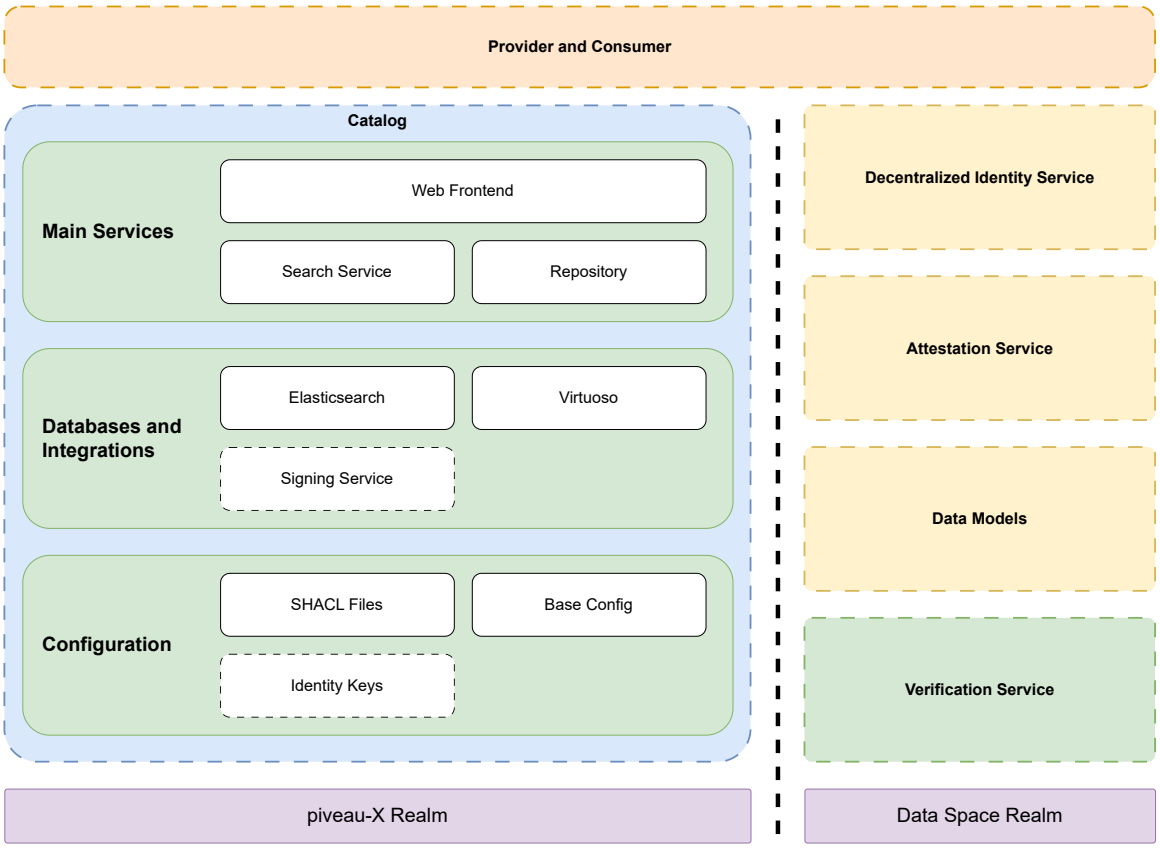


Figure 1: Service, components, and Data Space dependencies of piveau-X

Data Space Blueprint. As the foundation for our architecture, we chose to build upon an existing, mature catalog solution: piveau. While piveau is already a comprehensive and scalable data management catalog based on Semantic Web technologies, it can be extended to meet the requirements of a Data Space catalog.

4. Prototype

The piveau-X catalog instantiates this architecture in the Gaia-X Data Space domain[1]. In this case, the Data Space Realm consists of a Decentralized Identity Service that provisions and updates an organization's DIDs, DID Documents, and associated certificates used to identify organizations and for secure signing and verification of VCs [14]; an Attestation Service implemented through the GXDCH, which validates Legal Participant, Service Offering, and Data Resource VCs and issues compliance VCs [15]; Data Models based on Gaia-X Legal Participant, Service Offering, and Data Resource standards [16]; and a verification service, again leveraging the GXDCH to verify the integrity of these VCs.

In the Configuration layer, the piveau-X catalog leverages SHACL shapes to define custom metadata models across various domains. In this context, the Gaia-X Legal Participant, Service Offering, and Data Resource models are each represented as dedicated SHACL files.

Beyond defining the metadata schema, these SHACL files also specify how RDF metadata is indexed and stored as JSON in Elasticsearch. For example, the Gaia-X Core Ontology defines the attribute `gx:providedBy` for a Service Offering resource, indicating the Legal Participant responsible for providing the service [16]. The SHACL representation of `gx:providedBy` specifies a single resolvable IRI linking to the Legal Participant providing the service. Furthermore, this SHACL shape defines

[1]A demo can be accessed here: https://possible.fokus.fraunhofer.de

how `gx:providedBy` is indexed to JSON, using the piveau `StandardKeyword` indexing function and setting the indexed attribute name to `provided_by`. This dual representation of RDF and indexed JSON data enables the piveau-X catalog to support robust semantic querying via SPARQL and deliver efficient full-text search capabilities.

```
[
    sh:path gx:providedBy ;
    sh:name "Service Provider"@en;
    sh:description "A resolvable link to the participant self-description providing the
    ↪  service."@en;
    sh:minCount 1 ;
    sh:maxCount 1 ;
    sh:nodeKind sh:IRI ;
    sh:order 0 ;
    pv:mappingClass "StandardKeyword" ;
    pv:mappingName "provided_by" ;
],
```

Once the piveau-X catalog is configured, users register their organizations in the Gaia-X ecosystem as Legal Participants, with the resulting resources stored in the Databases and Integrations layer. To do so, users send an API request to the catalog whose payload conforms to the Legal Participant SHACL shape, which includes fields such as the organization's name, description, address, and Legal Registration Number (i.e., VAT ID). This information is then forwarded to the Gaia-X Notary Service, which returns a Legal Registration Number VC upon verifying the organization's legal existence. Next, a Legal Participant VC is composed and signed by a separate signing service developed specifically for the Gaia-X Data Space, using normalization and signing algorithms defined by Gaia-X and configured with private keys associated with the Legal Participant's corresponding DID [17]. Both VCs are sent to the GXDCH for compliance verification; if compliant, a Compliance VC is returned. Only upon receipt of this Compliance VC are the Legal Participant resource, Legal Participant VC, Legal Registration Number VC, and Compliance VC stored in the catalog, ensuring that any user can verify that the Legal Participant is indeed Gaia-X compliant. Notably, all resources are stored both as RDF in a Virtuoso Triplestore and as JSON in Elasticsearch, allowing for both SPARQL queries and performant search results.

As illustrated in Figure 2, Providers (Legal Participants offering services) publish Service Offerings by sending an API request conforming to the Service Offering SHACL shape. This payload includes fields like the Service Offering's name, description, contract policy, the providing organization, and associated Data Resources. A Service Offering VC and associated Data Resource VCs are created, signed, and submitted along with the organization's existing Legal Participant and Legal Registration Number VCs to the GXDCH. If compliant, the GXDCH issues a Compliance VC, after which the Service Offering resource and all related VCs are stored in the piveau-X catalog, maintaining Gaia-X compliance integrity.

In the Main Services layer, consumers can view catalog resources via a web-based interface that provides an overview of all Legal Participant and Service Offering resources, with Service Offerings further categorized based on whether they include aggregated Data Resources. From the Legal Participant details page, users can view the resource attributes and inspect the corresponding Legal Participant VCs, Legal Registration Number VC, and the Compliance VC returned by the GXDCH. Similarly, the Service Offering details page displays the Service Offering's attributes along with any associated Data Resources. Users can also view the corresponding Service Offering and Data Resource VCs and confirm their Gaia-X compliance via the Compliance VC. This interface thereby ensures transparency into the complete lifecycle of a resource—from creation to verification and storage.

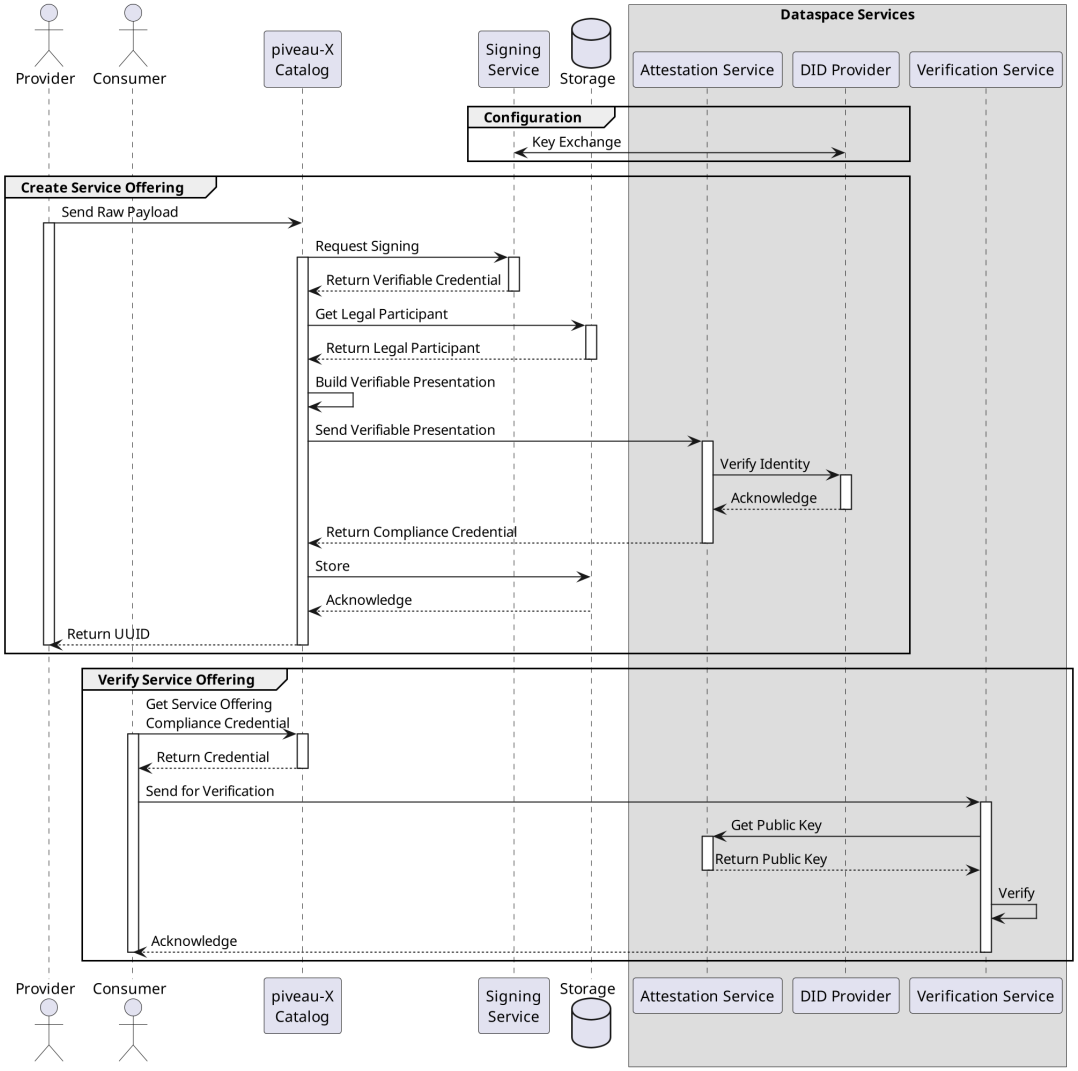


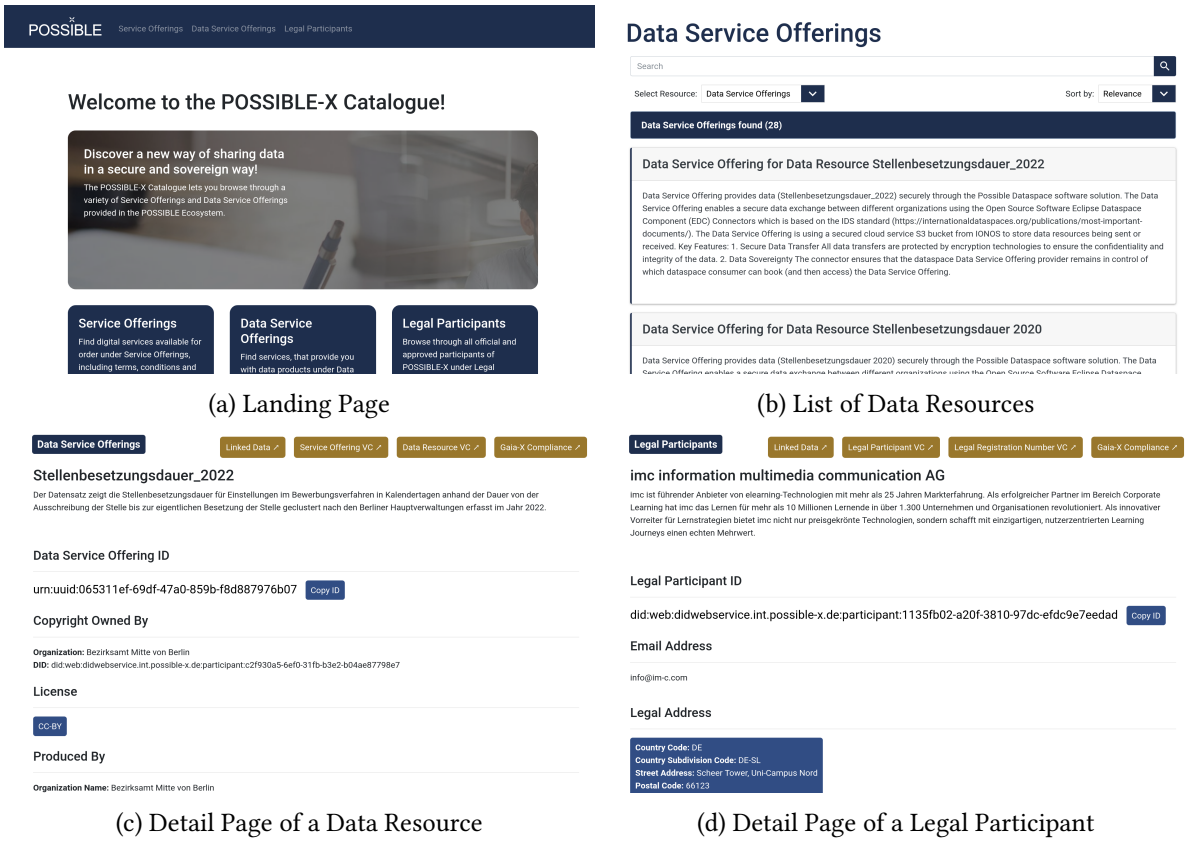
Figure 2: Service, components, and Data Space dependencies of piveau-X

5. Evaluation

We successfully validated piveau-X within the Gaia-X project POSSIBLE, whose objective is to demonstrate sovereign data value creation with real-world partners[2]. It integrates Gaia-X artifacts and other tools into a reusable toolbox to facilitate the bootstrapping and maintenance of a Data Space. Our solution is successfully integrated within the Data Space base layer of POSSIBLE, and is complemented by an end-user portal and the Eclipse Dataspace Components (EDC) Connector[3] for actual data exchange.

The following five pilot use cases leverage the POSSIBLE platform to provide Gaia-X-compliant Service Offerings and Data Resources: an interoperable messenger for public administration, sharing of non-public geodata, a skill analytics service, an AI-driven data product designer, a recommender for learning material, and a secure data-sharing service. For all stakeholders—both providers and consumers—of these use cases, we created and signed compliant Legal Participants in piveau-X, with the POSSIBLE portal acting as the gateway. Service Offerings are provided in a similar manner via the POSSIBLE portal and are cross-linked to the issuing participant within our catalog. POSSIBLE integrates with EDC connectors for data exchange and Open Digital Rights Language (ODRL) for describing policies. To manage the required information, such as EDC identifiers, we provided a SHACL-defined

[2]https://www.possible-gaia-x.eu/en/
[3]https://github.com/eclipse-edc/Connector



(a) Landing Page

(b) List of Data Resources

(c) Detail Page of a Data Resource

(d) Detail Page of a Legal Participant

Figure 3: Screenshots of the Prototype in the Scope of POSSIBLE

data model that seamlessly integrates into the GXDCH. The entire solution is deployed in a Kubernetes cloud, and the use case evaluation was successfully completed. Figure 3 shows some impressions of the user interface. Additionally, we developed a Verification Service[4] that wraps a real Gaia-X verification library to verify the technical compliance and correctness of signed artifacts.

Thus, the project has demonstrated that piveau-X can be effectively applied in real-world Data Space scenarios. Furthermore, the integration of Semantic Web technologies was instrumental in integrating the Gaia-X data model and accommodating use-case specific requirements. After all, the Gaia-X data model itself is based on SHACL and RDF.

6. Discussion and Future Work

In this paper, we presented our Data Space catalog implementation, piveau-X. Our solution enables the deployment and configuration of a catalog for a Data Space, allowing for the dynamic customization of the data model based on SHACL. As a result, the catalog can be tailored to the specific needs and requirements of a given Data Space. Furthermore, piveau-X provides direct integration points with identity and attestation services, making it a comprehensive solution for managing Data Space offerings. Furthermore, we successfully evaluated the catalog in real-world use cases within the Gaia-X research project.

In summary, our solution addresses challenges across all three technical building blocks of the DSSC blueprint. The direct use of Semantic Web technologies for defining data models enhances data interoperability. Integration with identity and attestation services ensures data sovereignty and trust. Finally, the core functionality that enables the discovery of offering descriptions allows data value creators to identify and reuse suitable data. Additionally, our approach aligns with the "Data Discovery"

[4] https://possible.fokus.fraunhofer.de/verifier/

challenge from the W3C Dataspaces Community Group, as the Semantic Web already provides a well-defined specification for describing, linking, and discovering data.

We are working to further extend and enhance piveau-X to reach production-grade maturity. piveau-X is a core component of a data trading platform developed in the scope of the European funded project PISTIS[5]. PISTIS emphasizes the monetization of proprietary data assets in a trustworthy and controlled environment, and is also following the core aspects of the DSSC blueprint. The project will demonstrate its outcomes in three large-scale demonstrators in mobility, energy and automotive domains. We are refining the generic data model design of piveau-X to enable even greater flexibility for use in PISTIS and beyond Gaia-X. PISTIS uses DLT for identity management and as a trust anchor. We are aiming to integrate this as another type of Attestation Service to demonstrate the pluggable trust mechanism of piveau-X.

Additionally, we are working to introduce extended lifecycle management for data, supporting multiple roles and states of offerings throughout the publication process. Finally, we intend to integrate a direct feature for data exchange based on the Data Space Protocol[6].

Acknowledgments

This work has been funded by the Federal Ministry for Economic Affairs and Climate Action (BMWK) under grant no. 68GX21009B (“POSSIBLE: Phoenix open software stack for interoperable engagement in dataspaces”) and by the European Union under the Grant Agreement 101093016 ("PISTIS: Promoting and Incentivising Federated, Trusted, and Fair Sharing and Trading of Interoperable Data ASsets").

Declaration on Generative AI

During the preparation of this work, the authors used *ChatGPT* solely for grammar and spelling checks. After applying the tool’s suggestions, the authors reviewed and edited the text and accept full responsibility for the published content.

References

[1] B. Otto, The Evolution of Data Spaces, in: B. Otto, M. ten Hompel, S. Wrobel (Eds.), Designing Data Spaces : The Ecosystem Approach to Competitive Advantage, Springer International Publishing, Cham, 2022, pp. 3–15. doi:10.1007/978-3-030-93975-5_1.

[2] Data Spaces Blueprint v1.5, 2025. URL: https://dssc.eu/space/bv15e/766061169/Data+Spaces+Blueprint+v1.5+-+Home.

[3] F. Kirstein, K. Stefanidis, B. Dittwald, S. Dutkowski, S. Urbanek, M. Hauswirth, Piveau: A Large-Scale Open Data Management Platform Based on Semantic Web Technologies, in: A. Harth, S. Kirrane, A.-C. Ngonga Ngomo, H. Paulheim, A. Rula, A. L. Gentile, P. Haase, M. Cochez (Eds.), The Semantic Web, Springer International Publishing, Cham, 2020, pp. 648–664. doi:10.1007/978-3-030-49461-2_38.

[4] Gaia-X Hub Germany, Gaia-X Explained, 2025. URL: https://gaia-x-hub.de/en/gaia-x-explained/.

[5] Gaia-X Technical Committee, Gaia-X Architecture Document, Technical Report 24.04, Gaia-X, 2024.

[6] W. W. W. C. (W3C), Decentralized Identifiers (DIDs) v1.0, 2022. URL: https://www.w3.org/TR/did-1.0/.

[7] W3C, Verifiable Credentials Data Model v2.0, 2025. URL: https://www.w3.org/TR/vc-data-model-2.0/.

[8] Data Spaces Support Centre, 2025. URL: https://dssc.eu/.

[5] https://www.pistis-project.eu/

[6] https://docs.internationaldataspaces.org/ids-knowledgebase/dataspace-protocol

[9] B. T. Arnold, K. Baydoun, D. Collarana, S. Duda, C. Gillmann, A. Hemid, P. Hertweck, P. Moosmann, D. Sukhoroslov, C. Lange, XFSC: A Catalogue of Trustable Semantic Metadata for Data Services and Providers, arXiv (2025). arXiv:2501.14473v1.
[10] GXFS-FR, Organisation and Partners, 2025. URL: https://www.gaia-x-hub.fr/en/organisation-and-partners/.
[11] Prometheus-X, Data Accessibility with the Prometheus Catalog Service, 2025. URL: https://prometheus-x.org/cbb02-catalog/.
[12] Prometheus-X, Prometheus-X Architecture, 2025. URL: https://prometheus-x.org/architecture/.
[13] Pontus-X, Open-Source Framework for the Industrial AI & Data Economy, 2025. URL: https://www.pontus-x.eu/.
[14] Dataport, Possible-x did web service, 2025. URL: https://github.com/Dataport/possible-x-did-web-service, accessed: March 7, 2025.
[15] Gaia-X Association, Digital Clearing House, 2025. URL: https://gaia-x.eu/services-deliverables/digital-clearing-house/, accessed: March 7, 2025.
[16] Gaia-X, Gaia-X Core Ontology, 2022. URL: https://gaia-x.gitlab.io/gaia-x-community/gaia-x-self-descriptions/core/core.html.
[17] G.-X. Lab, gaia-x-101, 2024. URL: https://gitlab.com/gaia-x/lab/workshops/gaia-x-101.

A. Online Resources

- Documentation: https://doc.piveau.io/hub/piveau-x
- Live Demo: https://possible.fokus.fraunhofer.de