Ref. Ares(2026)8298637 - 31/08/2026

AI-ENHANCED DATA UTILIZATION AND SMART CONTRACTING

Project deliverable D3.2

HORIZON Research and Innovation Action | 101203465 |
Call HORIZON-CL5-2023-D6-01

Funded by the European Union

Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

DELIVERABLE ADMINISTRATIVE INFORMATION

TITLE OF THE DELIVERABLE	D3.2 - AI-Enhanced Data Utilization and Smart Contracting
WORK PACKAGE	WP3
TYPE OF DELIVERABLE	R — Document, report
DISSEMINATION LEVEL	PU - Public
STATUS – VERSION, DATE	Final – V1, 24/08/2026
DELIVERABLE LEADER	AKKIS
CONTRACTUAL DATE OF DELIVERY	31/08/2026
SUBMISSION DATE	31/08/2026
KEYWORDS	Dataspace, Smart contracts, blockchain, Data utilisation, data discovery, AI, LLM, RAG, CKAN, Big Data, cloud, EMDS, anonymisation, cataloguing,

Funded by the European Union

Project funded by Schweizerische Eidgenossenschaft Confédération suisse Confederazione Svizzera Confederaziun svizra Swiss Confederation Federal Department of Economic Affairs, Education and Research EAER State Secretariat for Education, Research and Innovation SERI

LIST OF CONTRIBUTORS

Name	Organisation
Thierry Chevallier, Renée Obregon Gonzalez, Vincent Delmas, Jonathan Randriarison, Sylvain Pham	Akkodis
Kostas Cheliotis	FRONTIER Innovation

QUALITY CONTROL

	Name	Organisation	Date
Peer review	Julian Stephens	MJC2	3/08/2026
Peer review	Tania Panayiotou	UCY	7/08/2026
Quality check	Roozbeh Mohammadi	ERT	31/08/2026

VERSION HISTORY

Version	Date	Author	Summary of changes
0.1	22/07/2026	AKKIS, FRONT	Initial version
0.1_rev	14/08/2026	UCY, MJC2	Version reviewed
0.2	24/08/2026	AKKIS	Version for quality check
1.0	31/08/2026	ERT	Final version

LEGAL DISCLAIMER

Funded by the European Union

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

TABLE OF CONTENTS

Funded by the European Union

Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

Funded by the European Union

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

LIST OF FIGURES

Funded by the European Union

Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

LIST OF TABLES

Funded by the European Union

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

LIST OF ABBREVIATIONS AND ACRONYMS

Acronym	Meaning
API	Application Programming Interface
CKAN	Comprehensive Knowledge Archive Network
DCAT-AP	Data Catalogue Vocabulary Application Profile
DID	Decentralised Identifier
DKAN	Drupal-based Knowledge Archive Network
DSSC	Data Space Support Centre
EDC	Eclipse Dataspace Components
EMDS	European Mobility Data Space
IDSA	International Data Space Association
KB	Knowledge Base
KG	Knowledge Graph
LLM	Large Language Model
MDHESE	Mobility Data Harmonisation and Enrichment for Standard Exchange
MVD	Minimum Viable Dataspace
OLAP	Online Analytical Processing
RAG	Retrieval Augmented Generation
RM-ODP	Reference Model for Open Distributed Processing
UI	User Interface
VC	Verifiable Credential
VP	Verifiable Presentation
VRU	Vulnerable Road User
WP	Work Package

Funded by the European Union

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

PROJECT EXECUTIVE SUMMARY

Lack of orchestration, structured and standardized integration protocols and metadata descriptors, incorporation of real-world traffic complexities and nuances, underutilization of valuable resources, model uncertainties and integration of micro-mobility services and VRUs result in suboptimal performance in addressing complex issues related to the management of mobility services and infrastructure and a divergence from EU's sustainable mobility targets. FEDORA aims to pave the way towards advanced traffic and network management through the development of a federated spaces platform offering a holistic framework of innovative solutions and services that enable precise and pro-acting sensing of supply and demand, facilitate optimal operation of transport services, and advances learning and evolution in complex environments. At the operational level, FEDORA offers a collaborative space of data that can realize advanced data alchemy processes using interconnected services and tools, a space of advanced traffic management optimisation services and a multi-modal simulation space to create and assess future mobility scenarios. The approach is validated in six thematic demonstrations in Vienna, the Basque country, Reggio Emilia, Nicosia, Budapest and Denmark, covering varying EU urban and rural contexts, infrastructure maturity levels, multimodal mobility services availability, organisation/operational structures and social conditions. Interaction with existing programmes on road mapping and recommendations at national, EU and global level will be promoted, allowing a multiplication effect of project's results.

Social Media link:

@fedora-euproject

For further information please visit https://fedora-horizon.eu

Funded by the European Union

Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

DELIVERABLE EXECUTIVE SUMMARY

This deliverable presents the design and implementation of the FEDORA Data Space, with the aim to validate its vision of trusted, interoperable, and AI-enabled mobility data sharing. It showcases key technical components of the FEDORA Data Space and illustrates how they facilitate secure data exchange, efficient data discovery, and transparent governance.

The report presents three technical components and demonstrates their implementation. The first component complements the data catalogue presented in the D3.1 deliverable and consists of an **AI assistant enhancing the comprehension and usability of the data resources** made available through the public catalogue, supporting natural language interaction. The second component is the **Smart Contracts Layer**, which maintains a trusted and verifiable record of data-related transactions, supports the execution and verification of data-sharing agreements, and enhances transparency and trust among data space participants. The third component is the dataspace solution itself, i.e. the **Minimum Viable Dataspace** including dataspace connectors deployed in each partner organisation, that enable trusted and interoperable exchange of pilot site relevant data between data providers and consumers in accordance with common governance principles and interoperability standards.

The AI assistant is implemented using Dify, an open-source platform for building, orchestrating, and deploying LLM-powered applications and RAG workflows. Despite its strong feature set, Dify has some architectural limitations (related in particular to scheduling and data source integration), which required us to design a dedicated workflow to ingest CKAN metadata into the RAG knowledge base. The next steps are to enhance the RAG pipeline and evaluate a more agentic architecture.

The smart contracts layer uses Hyperledger Fabric, an open-source distributed ledger platform providing the required smart contract execution. The connection between the blockchain and the dataspace is realized via the so-called blockchain endpoint client, a standalone module deployed by each participant in the FEDORA Data Space, and more specifically via the smart contracts' listener. In a next step, the dataspace will make use of this infrastructure to ensure that data exchange operations in FEDORA are reflected on the blockchain, passing events as parameters.

The Minimum Viable Dataspace is built on Eclipse Dataspace Components (EDC), selected for its performance, decentralized identity support and alignment with mobility standards. Three technical partners formed a core MVD to validate EDC-based connectivity and basic data exchange. An initial configuration was tested through several iterations, highlighting the complexity of establishing a multi-partner dataspace environment. In the next iterations, decentralized trust services will enable participant and asset identification and authentication, and additional FEDORA partners will be included.

Overall, this report demonstrates the feasibility of integrating AI-enhanced data discovery services and trusted transaction management within a standardised mobility data space solution.

Funded by the European Union

Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

1 INTRODUCTION

The present deliverable (D3.2 – "AI-Enhanced Data Utilization and Smart Contracting") is a report accompanying the demonstration of the FEDORA Data Space (WP3), showcasing the practical data discovery, access and secure exchange between the project's stakeholders. It aims to illustrate the project's vision and includes: (i) a Minimum Viable Dataspace enabling secure and interoperable exchange of mobility-related data between partners, (ii) an AI-powered assistant that enhances the Public Data catalogue demonstrated in D3.1 by enabling natural language interaction and intelligent support for data discovery and access, and (iii) a Smart Contracts Layer that maintains a trusted and verifiable record of all data-related transactions between data space participants, while supporting the validation of transactions and the enforcement and verification of data-sharing agreements.

These components help achieve the project's impacts and goals. The Minimum Viable Dataspace supplemented by the Smart Contracts layer facilitate seamless and trusted data exchange between the stakeholders involved in multimodal traffic management (data providers, data users and systems). And the AI assistant provides a solution to data discovery challenges such as the limited understanding of data meaning and quality.

The report describes the objectives, architecture, implementation, and key functionalities of the corresponding components, as well as their integration into a coherent ecosystem. This integration is guided by a methodology outlined in the first section, which aims to relate the work presented in this deliverable to ongoing initiatives in the field of mobility data sharing (e.g. the European Mobility Data Space).

This report demonstrates the added value of combining a standardised data space with smart data agreements, as well as complementing it with AI-assisted services aiming to improve data discoverability, accessibility and reuse.

1.1 MAPPING FEDORA OUTPUTS

The Table 1 below show how this deliverable maps FEDORA Grant Agreement commitments, both within the formal Deliverable and Task description, against the project's respective outputs and work performed.

Table 1: Adherence to FEDORA GA Deliverable & Tasks Descriptions

FEDORA GA component title	FEDORA GA COMPONENT OUTLINE	RESPECTIVE DOCUMENT CHAPTER(S)	JUSTIFICATION
	DELIVERABLE		
D3.2 - AI-Enhanced Data Utilization and Smart Contracting	AI-driven data utilization tools and smart contract mechanisms for data sovereignty and sharing (T3.4, T3.5, T3.6)	Whole document	
	TASKS		
Task 3.4 – AI-facilitated data comprehension, utilization and validation	Enhances the FEDORA data catalogue with innovative AI assistant to avoid data	Chapter 3	Chapter 3 describes how the retrieval-Augmented Generation (RAG) technique

Funded by the European Union

Project funded by Schweizerische Eidgenossenschaft Confédération suisse Confederazione Svizzera Confederaziun svizra Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER State Secretariat for Education, Research and Innovation SERI

	miscommunication and data misunderstanding about what the datasets are and for what purpose they can be used.		is used for optimising the results of a large linguistic model (LLM) and how it is integrated within the FEDORA catalogue, as an extension (CKAN extension).
Task 3.5 – Smart contracts for data sovereignty and sharing	Establishes a decentralized governance framework for data sovereignty using blockchain technology, reducing the risks associated with centralized management of data access policies.	Chapter 4	Chapter 4 describes how the blockchain technology is used within the Data Space to immutably record all data-related transactions and embed data sharing agreement terms, timestamped and versioned, providing a clear, verifiable history of all changes.
Task 3.6 – Big Data enabled data space implementation	Implements the dataspace, facilitates seamless data exchange between the involved stakeholders, guarantees effective big data management, ensures that the system is capable of processing the wide variety of data sources for multimodal mobility and beyond, scalable, secure and interoperable, reuses as many existing building blocks of the EMDS as possible.	Chapter 5	Chapter 5 provides a detailed description of the dataspace components implemented in order to make the FEDORA Data Space compliant with the reference works of DSSC, IDSA/Gaia-X and European Mobility Data Space (EMDS).

1.2 DELIVERABLE OVERVIEW AND REPORT STRUCTURE

This deliverable, titled “AI-Enhanced Data Utilization and Smart Contracting,” primarily describes AI-driven data utilization tools and smart contract mechanisms for data sharing, which correspond respectively to tasks T3.4 and T3.5 of the FEDORA project workplan. Task T3.6, concerning the implementation of the dataspace and its relationship with the European Mobility Data Space (EMDS), although not explicitly mentioned in the title of this deliverable—is critical to the implementation of data sharing within the project and thus forms an essential part of this document. More specifically:

- Section 2 describes the methodology followed, introducing the RM-ODP viewpoints, the DSSC technical building blocks, the concept of Minimum Viable Dataspace (MVD), and the way to address the specificities of the FEDORA Data Space,
- Section 3 describes the AI-enhanced data discovery and utilisation, in particular how data discovery can be assisted by Large Language Models and Knowledge graphs, and how to integrate this assistant in a catalogue software system such as the FEDORA public catalogue,
- Section 4 describes the smart contracts for data sovereignty and sharing, with a particular focus on how to integrate the blockchain-based solution with the dataspace,

Funded by the European Union

Project funded by Schweizerische Eidgenossenschaft Confédération suisse Confederazione Svizzera Confederaziun svizra Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER State Secretariat for Education, Research and Innovation SERI

- Section 5 describes the data space implementation itself, with a description of the Data Space connector and the deployment of the Minimum Viable Dataspace
- Section 6 describes the software stack (technology viewpoint), going into the technological choices related to the metadata catalogue solution, the AI stack, the dataspace stack and the blockchain stack respectively,

Funded by the European Union

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

2 METHODOLOGY

This deliverable, namely "AI-Enhanced Data Utilization and Smart Contracting" is the counterpart of D3.1 "Integrated cross-sectoral Data Space".

Articulating seamlessly these two deliverables required us to carefully consider a methodology, as a purely functional breakdown (i.e. T3.1, T3.2, T3.3 on one hand, T3.4, T3.5, T3.6 on the other) did not prove fully adequate. Indeed, all WP3 tasks are closely interlinked in several ways: for instance, T3.1 defines technical requirements not only of the catalogue (T3.2) but also of the dataspace (T3.6). Another gap is the AI-enhanced data utilization (T3.4) which is tightly coupled to the Data discovery (T3.2).

It was therefore necessary to define an approach that allows the content to be divided into two distinct deliverables without this breakdown appearing arbitrary or uncomfortable for the reader. For this reason, the RM-ODP approach was followed for both deliverables, focusing on Enterprise and Information viewpoints in D3.1 and on Engineering and Technology viewpoints in D3.2.

From an engineering perspective, our work on the Data Space has been guided by a concept widely embraced in current data-sharing initiatives, namely the Minimum Viable Dataspace, which refers to the minimum configuration required to launch a functional data space. Guided by this proven concept, and having identified core technical requirements for the solution, we then focused on identifying the specific features of the data space to meet the unique needs of the pilot sites ("FEDORA specifics").

2.1 RM-ODP ENGINEERING AND TECHNOLOGY VIEWPOINTS

2.1.1 THE RM-ODP VIEWPOINTS

The RM-ODP[1] is an ISO/IEC standard that provides a structured framework for specifying complex distributed systems through a set of complementary viewpoints. As illustrated in Figure 1, it separates concerns into five perspectives—enterprise, information, computational, engineering, and technology—allowing stakeholders to describe objectives, data semantics, functional decomposition, infrastructure, and implementation choices in a consistent and interoperable manner.

[1] Reference Model of Open Distributed Processing

Funded by the European Union

Project funded by Schweizerische Eidgenossenschaft Confédération suisse Confederazione Svizzera Confederaziun svizra Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER State Secretariat for Education, Research and Innovation SERI

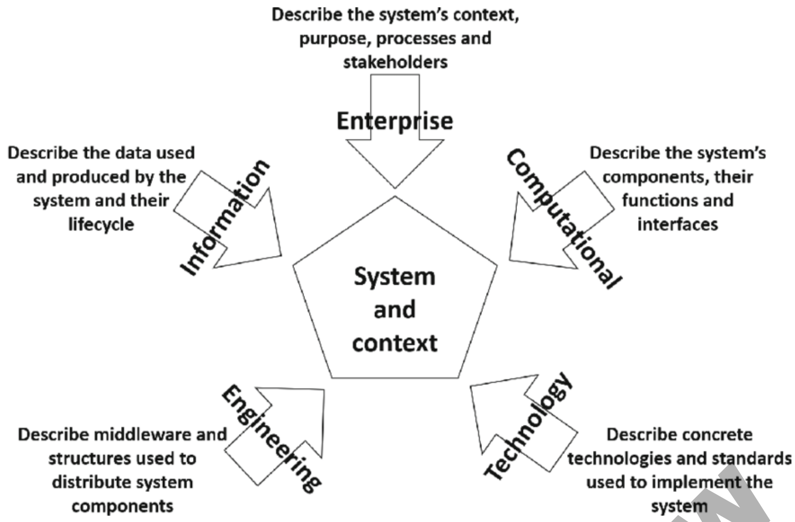

Figure 1: The five viewpoints of RM-ODP

Applying RM-ODP to the specification of the FEDORA dataspace, it provides a clear alignment between architectural design and specific requirements of the dataspace. More specifically it provides a methodological approach to organize heterogeneous actors, data flows, and services. **The enterprise viewpoint** formalizes governance, roles, and data-sharing agreements between public authorities, transport operators, and private service providers. **The information viewpoint** structures mobility data assets, including static supply data, dynamic flows, and metadata aligned with standards such as DCAT-AP. **The computational viewpoint** defines interoperable services for data exchange, discovery, and processing, which are essential for enabling e.g. multimodal journey planning, real-time analytics, dynamic pricing, network services optimisation, anomalies detection, etc. **The engineering viewpoint** addresses the distributed nature of the dataspace, ensuring secure, scalable, and sovereign data sharing across platforms. Finally, **the technology viewpoint** maps these specifications to concrete implementations, such as APIs, data formats, and cloud or edge infrastructures.

The Enterprise and Information viewpoints are fully addressed in the first half of the FEDORA Data Space specification (which is split into two deliverables), namely D3.1 "Integrated Cross-sectoral Data Space". This second half of the specification (namely D3.2), being more technical, focuses more on engineering and technology viewpoints. Both deliverables deal with the computational viewpoint on their respective specific components.

2.1.2 THE RM-ODP ENGINEERING VIEWPOINT ON THE FEDORA DATASPACE

The RM-ODP engineering viewpoint applied to the FEDORA dataspace describes it as a set of distributed components enabling secure and sovereign data exchange between participants, namely dataspace connectors. These connectors deployed within each participating domain typically include a data plane and a control plane, and support security and policy enforcement. The data plane is responsible for the actual transfer of data assets between providers and consumers, supporting both streaming and batch exchanges over standardized protocols. The control plane manages negotiation, contract establishment, and ensures that all transfers are governed by explicit usage policies.

The engineering viewpoint also describes the FEDORA data catalogue as a set of components for harvesting, transforming, synchronising, and accessing metadata across heterogeneous systems. It ensures interoperability with external data portals and dataspace connectors through multi-protocol harvesting and synchronisation services, supporting incremental updates, semantic alignment, and controlled metadata exchange. The catalogue is further enhanced with AI-driven components which include semantic

Funded by the European Union

Project funded by Schweizerische Eidgenossenschaft Confédération suisse Confederazione Svizzera Confederaziun svizra Swiss Confederation Federal Department of Economic Affairs, Education and Research EAER State Secretariat for Education, Research and Innovation SERI

enrichment, vectorisation, and RAG pipelines. Metadata is transformed into embeddings to support semantic search, while coordinated retrieval and generation workflows enable contextualised responses that improve data comprehension and usability.

2.1.3 THE RM-ODP TECHNOLOGY VIEWPOINT ON THE FEDORA DATASPACE

The RM-ODP technology viewpoint applied to FEDORA specifies the concrete technology used to implement the data exchanges in the dataspace, namely the Eclipse dataspace components. The technology viewpoint also defines how the data space integrates the CKAN catalogue software solutions and metadata standards enabling discovery through REST APIs and alignment with DCAT-compatible metadata models such as mobilityDCAT-AP. The technology viewpoint also specifies the concrete technologies used for the deployment, e.g. Docker images for the connector, orchestrated via Kubernetes or Nomad.

2.2 THE DATA SPACE TECHNICAL BUILDING BLOCKS

From a technical building-block perspective, the DSSC outlines three primary categories[2]: data interoperability, data sovereignty and trust, and enablers for data value creation (Figure 2).

Figure 2: Overview of the DSSC technical building blocks (source: DSSC blueprint)

- **The Data interoperability building block** ensures that data remains usable, reliable, and operational across the entire data space. More specifically:

Data models ensure that data produced by one participant can be correctly interpreted by another without additional transformation, by relying on shared schemas, controlled vocabularies, or ontologies. In FEDORA, the “dataset harmonisation, enrichment, and synthetisation” methods (Task 3.3, more details in D3.1) can be considered as an implementation process of the data models within the data interoperability building block, as it enables consistent, meaningful, and reusable data across the dataspace. To a lesser extent, this task also relates to data exchange and metadata/semantic alignment. The support for mobilityDCAT-AP in “Data discovery, cataloguing and publication” (Task 3.2) strengthens the data models / metadata alignment aspect of data interoperability. By using a standardized metadata profile, the catalogue ensures that mobility datasets are described in a consistent, machine-readable way, enabling cross-domain and cross-platform understanding.

[2] DSSC blueprint https://blueprint.dssc.eu/?pane=technical

Funded by the European Union
Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

Data exchange mechanisms provide standardized interfaces and protocols through which data is exposed and consumed. In FEDORA, this corresponds to the “Big data enabled data space implementation” (Task 3.6) in which distributed storage systems, real-time ingestion, and stream processing are fundamentally about enabling scalable and continuous data exchange. Stream processing frameworks and real-time pipelines ensure that data is not only exchanged but made available with the required velocity and granularity. While DSSC does not define “big data infrastructure” as a standalone building block, it is implicitly part of the technical enablers supporting interoperable data exchange mechanisms.

Provenance and traceability ensure that data consumers can assess compliance with usage policies, which is critical in multi-actor environments where data is reused and combined across organizational boundaries. **Observability** enables participants to track how data is accessed, processed, and shared in real time, through logging, monitoring, and auditing capabilities. In FEDORA this building block corresponds to “Smart contracts for data sovereignty and sharing” (Task 3.5) and its blockchain-based implementation.

- **The Data sovereignty and trust** building block ensures that data sharing remains controlled by the data provider’s rules while still enabling interoperability and scalable collaboration across organizational boundaries.

In FEDORA, the data sovereignty and trust layer matches primarily the blockchain-based decentralized governance framework for data sovereignty (Task 3.5). More specifically, smart contracts map to **usage policy enforcement** in the DSSC sense. They allow data access and usage rules to be encoded as executable logic, so conditions like “who can access what, under which constraints” are enforced automatically and transparently. This aligns closely with the DSSC notion of sovereign data control, where governance is not only declarative but technically enforceable across organizational boundaries. The blockchain layer may contribute to the **trust framework**. Instead of relying on a central authority to manage identities, validate participants, and enforce rules, the governance logic is distributed across a ledger. The blockchain-based mechanisms can also reinforce **identity management**, particularly when combined with decentralized identifiers and verifiable credentials, by enabling self-sovereign identity models that remove dependency on centralized identity providers.

- **Data value creation enablers** aim to transform governed data into consumable assets and services that can generate economic or operational value.

In FEDORA, the **publication and discovery** component primarily matches the “Data discovery, cataloguing and publication” (Task 3.2). Indeed, a public catalogue exposes data offerings so they can be found, accessed, and reused across the ecosystem. The harvesting functionality extends this role by enabling automated ingestion of external datasets, which effectively turns the catalogue into a federated entry point for multiple data sources, including other data spaces. This directly supports the DSSC objective of scalable discovery across distributed environments rather than isolated repositories.

The **publication & discovery** and **data services & offering descriptions** also correspond to the “AI-facilitated data comprehension, utilisation and validation” (Task 3.4). Indeed, an AI assistant integrated into a public-facing catalogue enhances how data offerings are exposed and understood. By clarifying dataset meaning, intended use, and context, it directly augments the offering descriptions, making them more accessible and less prone to misinterpretation. In parallel, it improves the discovery process, not by changing the underlying metadata standards (e.g., DCAT-AP), but by adding an intelligent interaction layer that helps users navigate, interpret, and select relevant data assets more effectively. This type of assistant can also be seen as a value creation service in its own right. It consumes catalogue metadata and possibly underlying datasets to generate explanations, recommendations, or usage guidance, thereby increasing the practical usability and adoption of the data space.

Funded by the European Union

Project funded by Schweizerische Eidgenossenschaft Confédération suisse Confederazione Svizzera Confederaziun svizra Swiss Confederation Federal Department of Economic Affairs, Education and Research EAER State Secretariat for Education, Research and Innovation SERI

In FEDORA, **Value creation services** can be mapped, at least partially, with some features "Big-data enabled data space implementation" (Task 3.6). Real-time analytics on traffic conditions or transit data transforms raw data streams into actionable insights, which is precisely the role of value creation services operating on top of interoperable data flows.

The

Table 1 below provides a mapping between the technical components of the FEDORA Data Space and the corresponding DSSC building blocks (primary and secondary).

Table 2: Mapping between FEDORA components and DSSC building blocks

FEDORA component	Primary DSSC technical building block	Secondary DSSC building block	Rationale
Data discovery, cataloguing and publication	Data value creation enablers (Publication & discovery)	Data interoperability (Data- or metadata - models)	Provides federated dataset discovery and ingestion with standardized semantic description (mobilityDCAT-AP)
Dataset harmonisation, enrichment, and synthetization for complex data exchanges	Data interoperability (Data models)	Data exchange	Ensures semantic + syntactic consistency for cross-system usability
AI-facilitated data comprehension, utilization and validation	Data value creation enablers (Publication & discovery)	Data services & offering descriptions	Adds semantic guidance layer improving interpretation and selection of data offerings
Smart contracts for data sovereignty and sharing	Data sovereignty & trust (Usage policy enforcement)	Trust framework	Encodes and automates enforceable data usage conditions
Big-data enabled data space implementation	Data interoperability (Data exchange)	Value creation services	Enables high-volume, high-velocity data flows and real-time analytics pipelines

2.3 THE MINIMUM VIABLE DATASPACE

The concept of MVD originates from the work of the International Data Spaces Association (IDSA), which proposed a practical way to bootstrap a dataspace with the minimum set of components required for secure and sovereign data exchange. According to IDSA, an MVD consists of "just enough features to be usable for secure and sovereign data exchange"[3] and is intended as a starting point that can evolve iteratively. In other words, the MVD can be seen as the smallest functional implementation of a dataspace that

[3] https://dssc.eu/space/SK/812515476/11+-+Appendix+2:+Resource+Inventories

Funded by the European Union

Project funded by Schweizerische Eidgenossenschaft Confédération suisse Confederazione Svizzera Confederaziun svizra Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER State Secretariat for Education, Research and Innovation SERI

demonstrates the feasibility and value of trusted data sharing for a specific use case while relying only on essential technical and governance components.

The MVD can be understood as the minimal implementation profile of the DSSC technical building blocks that is sufficient to enable operational cross-organizational data sharing. The MVD operationalizes the DSSC framework by selecting only the essential building blocks, and typically includes participant onboarding, secure data exchange, policy enforcement, and metadata publication and discovery. The MVD approach allows an incremental adoption mechanism enabling the FEDORA consortium to deploy an interoperable data space with reduced implementation complexity (while remaining fully aligned with the DSSC architecture). More advanced capabilities, such as semantic interoperability or more complex value-added services, are intentionally deferred to later stages of the project, as the Data Space maturity increases.

2.4 FEDORA SPECIFICS

The concept of Data Space in FEDORA, as one of three integrated spaces (alongside a service space and a simulation space), goes beyond the dataspace definition established by the DSSC by incorporating additional functionalities that turn it into a broader data platform. Dataset harmonization, enrichment and synthesis, AI-facilitated data comprehension, utilization and validation, and the blockchain-based governance framework complement the nominal capabilities of the dataspace as defined by the reference dataspace architectures (IDSA, Gaia-X, etc.).

2.4.1 KEY TAKEAWAYS FROM THE FRAMEWORK ARCHITECTURE

The D2.2 deliverable ("Living Labs Setup, Spaces Definition, and Evaluation Models") does not only describe the Living Lab methodology for the pilot sites, it also presents the conceptual framework architecture for the so-called FEDORA spaces. More specifically it defines these three spaces (the data space, the optimisation services space and the simulation space) and their interoperability principles. It therefore serves as an indispensable reference for the present document. In this section, we outline the key findings from this report and the technical requirements that enable us to implement the Data Space solution.

2.4.1.1 FEDORA SPACES DEFINITION

According to the conceptual framework architecture, FEDORA categorizes its technical work into three distinct Spaces, each containing specific tools, solutions and components with distinct functionality and aims. The three Spaces of the FEDORA project are the Data Space, the Network Management Services Space, and the Simulations Space.

The Data Space, in this understanding, aims to facilitate secure and trusted data sharing within the FEDORA platform. All data sharing within FEDORA should take place within it. The workflow of the Data Space is to receive datasets from relevant trusted sources, to enrich/harmonize them, and to make them available to the various FEDORA solutions. Consequently, it takes as input raw data from external sources and outputs harmonized and enriched data.

The other two spaces each take data as input for distinct processing operations (respectively the generation of traffic network optimization solutions and the provision of simulation testbeds for exploring scenarios and solutions) – data which may correspond to the output data from the Data Space.

Funded by the European Union

Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

2.4.1.2 COMPONENT SPECIFICATIONS

Each of the FEDORA spaces is made up of components, defined by their inputs, outputs, processes, and characteristics. For the Data Space, the components identified and defined in the framework architecture are the following:

1. FEDORA Public Catalogue,
2. Mobility Data Harmonisation, Enrichment & Synthetisation Engine,
3. RAG Search System,
4. Smart Data Contracts Module,
5. Dataspace Connector

The first two components were described in greater detail in D3.1. The next three (**RAG search system**, **Smart Data Contracts Module**, and **Dataspace Connector**) are covered in the following sections of this document.

2.4.2 SPECIFIC ASPECTS OF THE FEDORA DATA SPACE

We can observe a certain difference between the definition of the Data Space in the FEDORA framework architecture and the definition of the data space according to the DSSC. Indeed, the DSSC defines a data space as a decentralized ecosystem that enables trusted data sharing and use among participants. In this sense, a dataspace is mainly concerned with common governance, interoperable data exchange standards, and sovereign control over data. The DSSC definition thus focuses on the governance and infrastructure that enable data sharing, not on the data preparation activities themselves. This means that data and metadata harvesting, data enrichment and harmonisation, and other data preparation activities, defined as key functionalities of the FEDORA Data Space, are not *defining* properties of a dataspace in the usual DSSC sense - although they may *implicitly* be part of it.

What does this mean? – That only specific tasks in the FEDORA Data Space deal with the implementation of nominal dataspace capabilities, while the others deal with additional capabilities, that are key to the "Data Space" in the FEDORA sense and can be supported by this Minimum Viable Dataspace.

The task that specifically addresses the dataspace's compliance with the DSSC building blocks in general, and with their MVD subset in particular, is the Task T3.6 "Big Data enabled dataspace implementation", which corresponds to section 5 of the present document, namely "Data space implementation". This Task is critical for the implementation of the Minimum Viable Dataspace.

2.4.2.1 DATA CATALOGUING

When it comes to nominal and additional Data Space capabilities, data cataloguing is of particular interest since, in the FEDORA sense, it encompasses both the nominal cataloguing of the dataspace (as listed among the building blocks of the Minimum Viable Dataspace, cf. supra) and a broader form of cataloguing, enabling an extensive collection of data across all categories of data sources (the public catalogue component identified in the framework architecture and described in greater detail in D3.1).

It is important to highlight here the significant contribution of FEDORA to this topic, as the data ecosystem in transport and mobility in general and in multimodal traffic management in particular is highly heterogeneous in terms of types, formats, access conditions, openness, and/or confidentiality. This heterogeneity particularly manifests itself through three main categories of data sources, namely Open data portals, Proprietary APIs and Dataspaces. Open data portals can be catalogued and exposed through portals such as CKAN offering easy discovery and immediate accessibility. Proprietary APIs, such as HERE services, are external services that can be listed in the catalogue as documented assets, including endpoints, service types, formats, licenses, and usage restrictions. Dataspaces per se are designed for the secure sharing of sensitive or strategic data, with fine-grained access control, traceability, and multi-stakeholder interoperability.

Funded by the European Union

Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

In this context, the cataloguing task does not come without ambiguity. Listing open datasets, public feeds, and documented APIs to ensure easy discovery for all users is not the same as listing sensitive assets that are subject to very strict access controls, governance policies, and full traceability via e.g. dataspaces. FEDORA adopts a dual approach, proposing two levels of cataloguing: the first level, which is public and centralized, enables comprehensive discovery of multimodal traffic management data both within and outside the FEDORA dataspace; and the second level, which is federated at the dataspace connector level, enables peer-to-peer discovery of conditional data offers between dataspace members. This dual approach will be described in greater detail in section 5.

2.4.2.2 SMART CONTRACTS

The same applies to Smart contracts (Task 3.5 and section 4 of the present document) which are not part of the nominal (core) dataspace capabilities but are a key feature of the FEDORA Data Space. Smart contracts, using blockchain as an implementation technology, complement the nominal dataspace capabilities in order to automate aspects of dataspace governance.

Funded by the European Union

Project funded by Schweizerische Eidgenossenschaft Confédération suisse Confederazione Svizzera Confederaziun svizra Swiss Confederation Federal Department of Economic Affairs, Education and Research EAER State Secretariat for Education, Research and Innovation SERI

3 AI-ENHANCED DATA DISCOVERY AND UTILISATION

FEDORA's research has highlighted the current limitations of data catalogues, particularly in the mobility domain, where users often misinterpret datasets due to ambiguous descriptions, overly technical or insufficient metadata, and a lack of explicit linkage to use cases. Such gaps can lead to incorrect interpretations, inappropriate use, or underutilization of data, posing a significant barrier to developing value-added services in a dataspace context. To address these challenges, several complementary approaches have been identified, including systematic annotation of datasets by use case, enriching metadata with explanations accessible to non-experts, and providing information on limitations, biases, and conditions of use.

Advanced LLM techniques, particularly RAG, appear as a relevant solution to transform a static catalogue into an interactive system that supports understanding and proper use of datasets. RAG combines text generation capabilities with dynamic access to external information sources, such as metadata, documentation, or usage examples, enabling the production of contextually relevant responses tailored to user needs.

In addition to RAG, other approaches have been explored, such as Knowledge Graphs (KGs) which provide advantages in explainability, robustness, and governance, and can be combined in a hybrid architecture to leverage both semantic formalization and the flexibility of language models.

3.1 DATA DISCOVERY ASSISTED BY LARGE LANGUAGE MODELS AND KNOWLEDGE GRAPHS

The objective is to provide users with a tool capable of addressing questions related to intermodal coordination, congestion forecasting, route optimization, and the assessment of incident impacts across different modes of transport. In this context, the proposed approach combines a structured KG with an LLM to integrate the reliability of structured data with the interpretive capabilities of natural language understanding.

Traditional keyword searches in catalogue software systems such as CKAN often fall short when addressing specific or layered questions, such as "Can you provide a descriptive analysis of dataset X?", "Which stakeholders have access to this data?", or "Does this dataset specifically include subway line timetables?" Using keyword search capabilities for these queries is often limiting as it requires users to perform multiple iterative searches, even when they know precisely what they are looking for.

The FEDORA AI-enhanced solution described in this section is designed to provide direct answers and detailed summaries. This solution is also capable of answering contextual questions related to the FEDORA project's content through an integrated project Knowledge Base (KB), a capability that is not provided natively by CKAN catalogues or comparable cataloguing system.

3.1.1 PROPOSED APPROACH

The proposed architecture (Figure 3) relies on orchestration between the KG, the LLM, and, optionally, a text-based search index for approximate queries. The KG contains entities representing datasets, transport modes, network segments, multimodal hubs, and incidents, as well as relationships such as bus lines traversing road segments or incidents localized to specific segments. The LLM interprets natural language queries, generates formal queries, synthesizes results, and can propose KG enrichment by analysing new data sources. This approach combines the rigor of a semantic model with the flexibility and contextual intelligence of an LLM.

Funded by the European Union

Project funded by Schweizerische Eidgenossenschaft Confédération suisse Confederazione Svizzera Confederaziun svizra Swiss Confederation Federal Department of Economic Affairs, Education and Research EAER State Secretariat for Education, Research and Innovation SERI

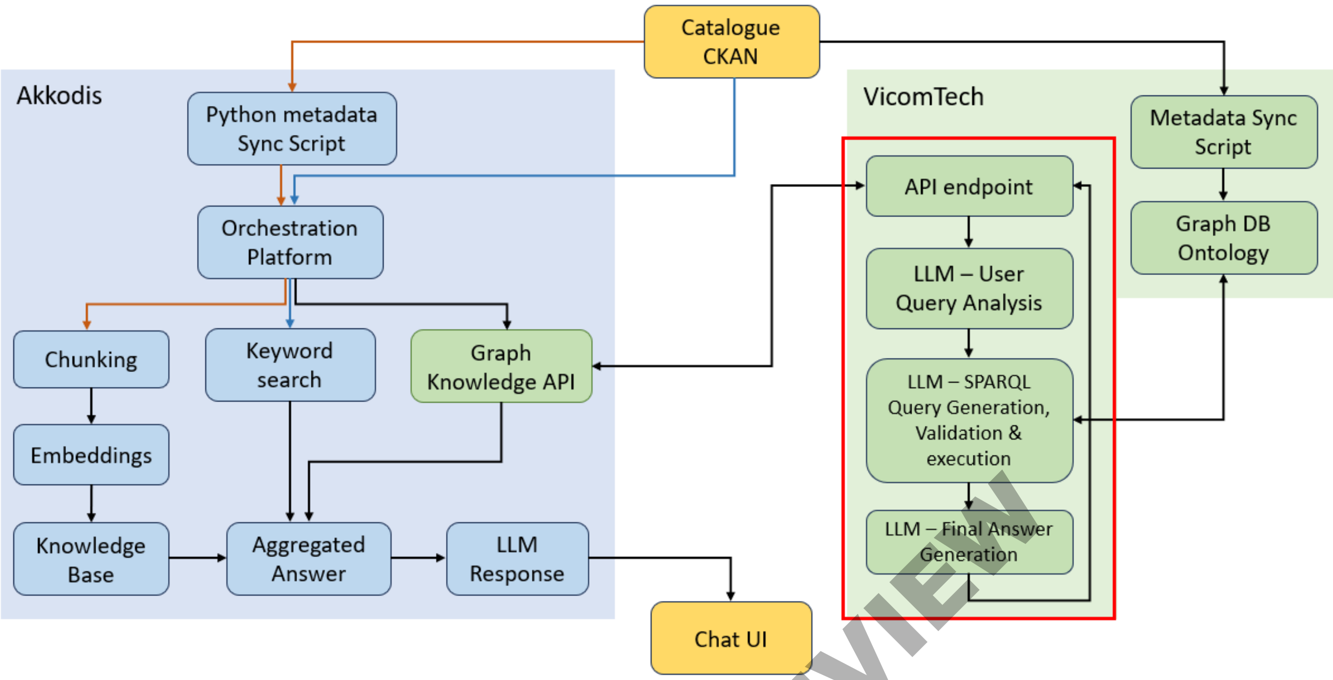

Figure 3: Orchestration between the Large Language Model and the Knowledge Graph

This combination offers many advantages. The KG ensures precision, interoperability, and semantic reasoning, while the LLM facilitates understanding complex questions and producing synthesized responses. The search index, when used, improves rapid retrieval and tolerance to linguistic variation. Limitations include the need to maintain a coherent ontology, validate LLM-proposed enrichments, and mitigate the risk of hallucinations when the model attempts to infer missing information.

3.1.2 IMPLEMENTATION

From a technical standpoint, it has been decided to build the AI platform on the open-source framework Dify (more details in section 6.2). The Dify framework is self-hosted on an Azure Virtual Machine, ensuring full control over the underlying infrastructure and its management. A dedicated data pipeline automates the extraction of metadata from the CKAN catalogue into our vector database (Knowledge Base).

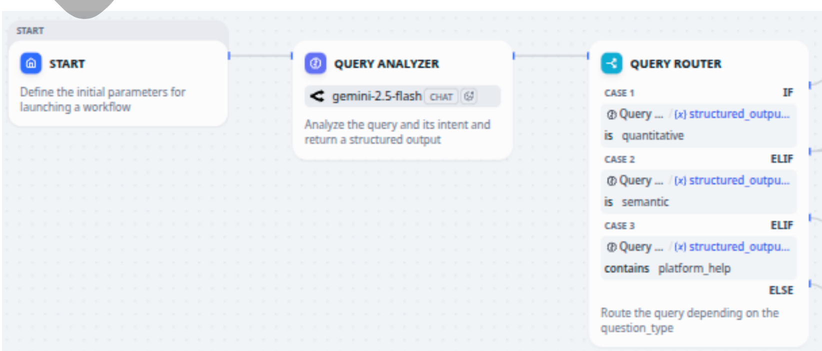

Figure 4: RAG pipeline (part 1)

Funded by the European Union

Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

To manage the potential complexity of the queries, the RAG orchestration uses a structured workflow with a semantic “query analyser” rather than a fully autonomous agentic approach (Figure 4). This query analyser evaluates the user's intent and directs the request through specific conditional pipelines optimizing speed, accuracy, control and telemetry (Figure 5).

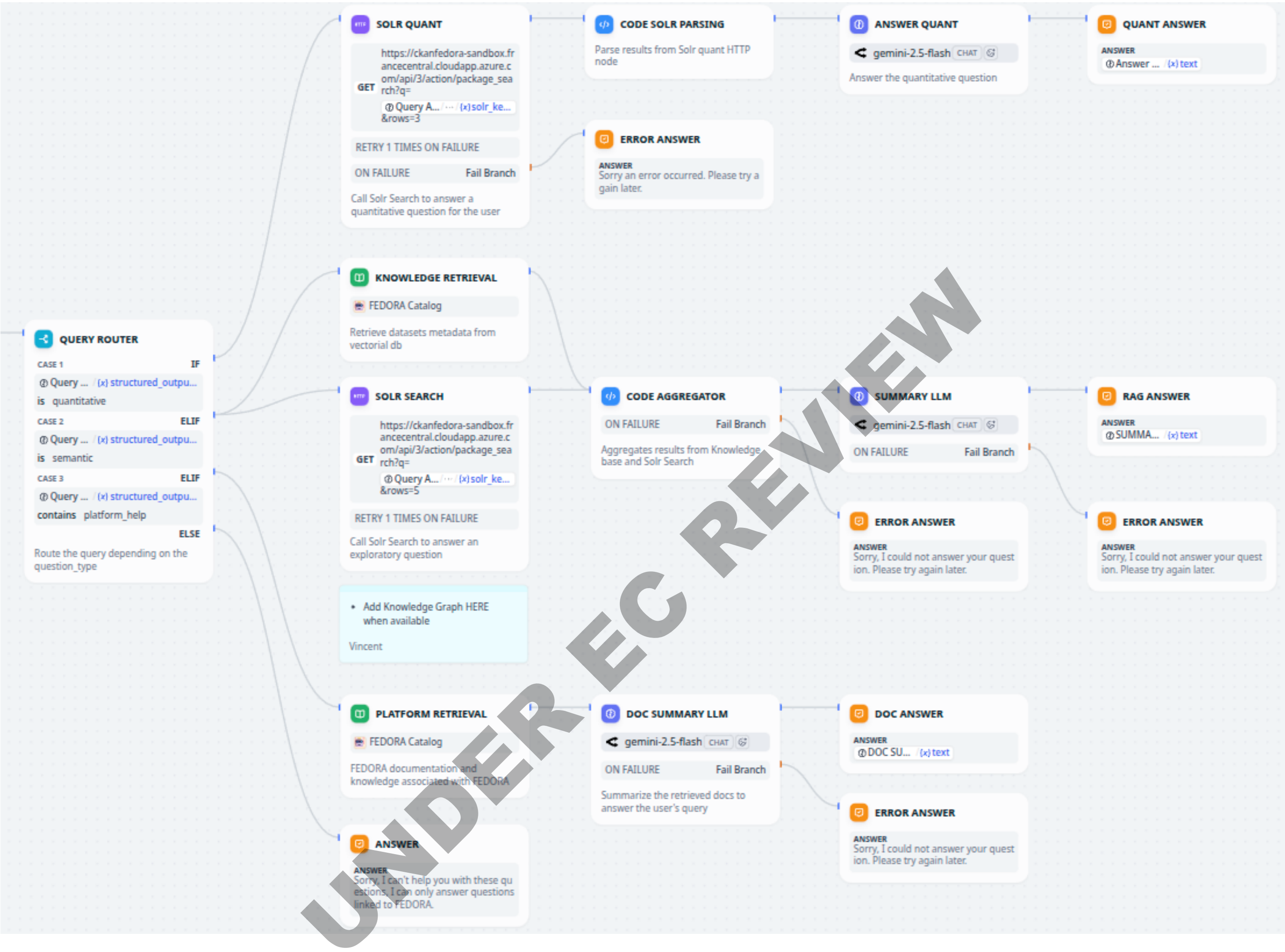

Figure 5: RAG pipeline (part 2)

Nevertheless, we still plan to evaluate a fully agentic reasoning RAG system to assess the strength and weaknesses of each approach, potentially creating a meta-analyser which, depending on the question could either launch the agentic pipeline or the workflow pipeline.

3.2 INTEGRATION WITH THE PUBLIC CATALOGUE

3.2.1 METADATA SYNCHRONISATION

CKAN metadata synchronization is implemented through an automated workflow consisting of three main stages: a scheduled trigger, the synchronization of CKAN content with the Dify Knowledge Base, and the generation of the resulting output. This workflow is illustrated in Figure 6 below.

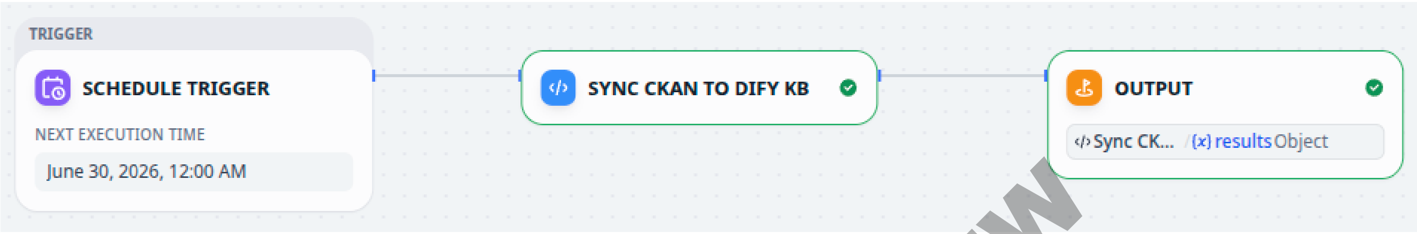

Figure 6: RAG CKAN metadata synchronisation

The central synchronization component contains the Python-based logic responsible for managing the KB content, including the creation, updating, and deletion of documents to ensure that the information stored in Dify remains aligned with the CKAN metadata.

A scheduler is configured to automatically trigger this synchronization process on a weekly basis, enabling regular updates of the KB without requiring manual intervention.

3.2.2 DATA ACCESS AND GOVERNANCE

Regarding the data access and governance, the AI system works securely by only querying the freely accessible metadata from the CKAN catalogue, without needing user authentication. To provide a deeper, file-level context, the LLM can also query the CKAN Datastore API for open-source datasets, allowing it to answer column-level queries. However, for private datasets, the system restricts the LLM to only search the metadata, warranting privacy and security.

3.2.3 USER INTERFACE

At the user interface (UI) level (Figure 7), the system currently uses an external chatbot web application that processes dataset links, retains conversation history within threads, saves data locally (browser local storage), and allows for the regeneration of responses.

Funded by the European Union

Project funded by Schweizerische Eidgenossenschaft Confédération suisse Confederazione Svizzera Confederaziun svizra Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER State Secretariat for Education, Research and Innovation SERI

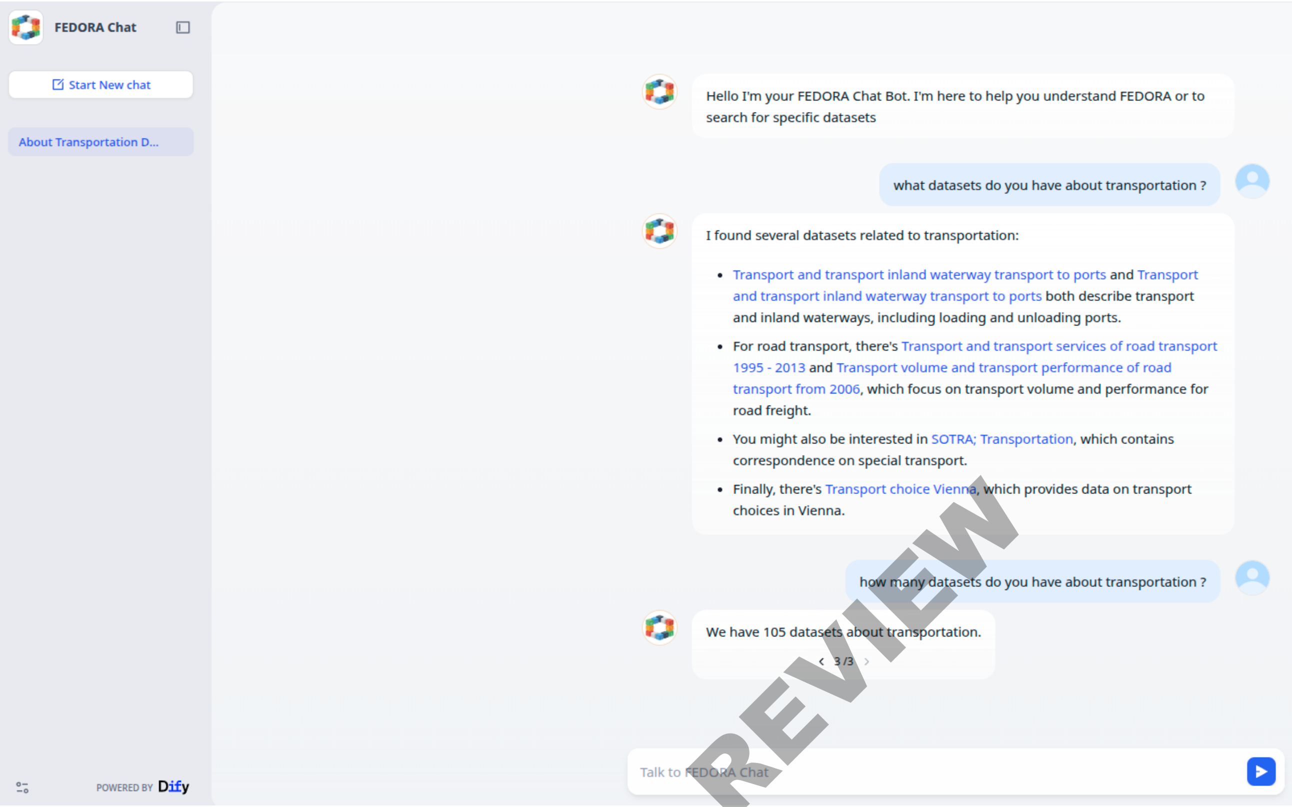

Figure 7: AI assistant chatbot interface

Moving forward, the objective is to embed this UI into the CKAN catalogue directly. This could be done either by redirecting the native search bar to the AI chat or by embedding a dedicated chat interface and floating widget directly within the CKAN portal to provide a unified, seamless user experience.

3.3 NEXT STEPS

The introduction of RAG into data catalogues represents a shift from traditional search paradigms toward more interactive and context-aware data exploration. While search engines based on technologies such as Solr provide efficient indexing and retrieval capabilities, RAG systems can generate synthesized responses and insights across multiple datasets. However, the added value of RAG must be assessed against its complexity, resource requirements, and potential risks related to accuracy and explainability. Determining whether RAG significantly enhances user experience and decision-making in the context of data catalogues requires empirical evaluation.

To measure this added value, the next phase of development involves creating evaluation metrics. We will aggregate a "golden dataset" of realistic Questions & Answers both crowdsourced from the FEDORA consortium and artificially generated by the Akkodis team. This baseline will be evaluated against multiple performance metrics to systematically assess the response quality, accuracy, and retrieval relevance. Using this feedback loop will allow for better fine-tuning of the whole workflow (query analyser, aggregation layer, etc.) and the LLM models, to make sure that the final AI system is robust, precise, and genuinely useful.

Funded by the European Union

Project funded by Schweizerische Eidgenossenschaft Confédération suisse Confederazione Svizzera Confederaziun svizra Swiss Confederation Federal Department of Economic Affairs, Education and Research EAER State Secretariat for Education, Research and Innovation SERI

4 SMART CONTRACTS FOR DATA SOVEREIGNTY AND SHARING

In the FEDORA dataspace, data sovereignty and trusted sharing are handled by the inclusion of an additional layer, that of the smart contracts. This smart contracts' layer needs to be able to fulfil the following key objectives: first, to keep a trusted record of all data-related transactions taking place between dataspace participants; and second, to allow dataspace participants to verify transactions and data-sharing agreements. In order to fulfil these objectives, a framework needs to be introduced which is both interconnected to some dataspace processes while at the same time operating independently. Furthermore, this layer has the additional requirement that any information recorded within it is done so in an immutable manner, meaning that information may be added, but not removed or changed.

This setup provides the following benefits: First, its interconnectedness to the dataspace will ensure that operations that take place in the dataspace are also automatically recorded in this independent layer; second, its immutability characteristic will ensure that all information recorded within has not been tampered; and third, due to its independent operation, dataspace participants can verify any dataspace transaction against this layer, that acts as an auditor to the dataspace.

Furthermore, this setup addresses all four requirements identified in the Grant Agreement regarding Task 3.5: Smart contracts for data sovereignty and sharing: The first and second requirements to "immutably record all data-related transactions in the form of meta-data" and "enable [cryptographical] verification of data authenticity and integrity" are addressed via both the immutability characteristic and the interconnectedness to the dataspace, ensuring that all data-related transactions are recorded immutably and can thus be verified at any point in the future. The third requirement to "embed mutually trusted and tamper-proof data sharing agreement terms, timestamped and versioned, providing a clear, verifiable history of all changes" is supported by the above properties and also the third property of functioning as an independent auditor to the dataspace, as data-sharing agreements constitute operations within the dataspace and any such agreement formed will also be stored in the independent layer. Finally the fourth requirement to "reply to enquiries on data accessibility according to said terms" is addressed by the auditor property, with the addition of a queries interface that will allow participants to perform queries on the immutably stored information.

4.1 BLOCKCHAIN FOR DATA SOVEREIGNTY AND SHARING

In order to realize the independent system mentioned above, we propose the introduction of blockchain technology. More specifically, we propose a solution where all dataspace partners are automatically enrolled within a blockchain network in addition to being connected via the dataspace. This introduces the concept of the blockchain ledger, which is an immutable record that contains all transactions within a network, and which all network participants have a local copy of. Any time a new transaction takes place it is automatically added to the ledger and transmitted to all participants, so each partner always has a verified copy of the latest changes. Furthermore, we propose the introduction of smart contracts coupled to the dataspace connectors, so that any dataspace transaction is automatically recorded in the blockchain via the smart contract.

This setup inherently addresses all the requirements outlined above. The interconnectedness with dataspace operations is handled via the smart contracts, which ensures that any dataspace operation (e.g. data sharing agreements, data transfers) is immediately transmitted to the blockchain. The immutability of records is done inherently via the use of a blockchain, whose core characteristics ensure that no record may be deleted. Finally, the fact that all participants also have a copy of the ledger means that any one participant can at any point verify the authenticity of any transaction.

Funded by the European Union

Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

4.2 BLOCKCHAIN-DATASPACE INTEGRATION

We present here the conceptual architecture required for the blockchain framework to support the operations outlined above. First, we introduce the **blockchain endpoint client**, which is the top-level module that contains within it all of the sub-components that handle the various operations. This blockchain endpoint client will be provided as a standalone module, and is expected to be deployed at any system that is also a FEDORA dataspace participant, to ensure that all dataspace operations are also reflected in the blockchain. A high-level outline of the connection between the dataspace and the blockchain is presented in Figure 8.

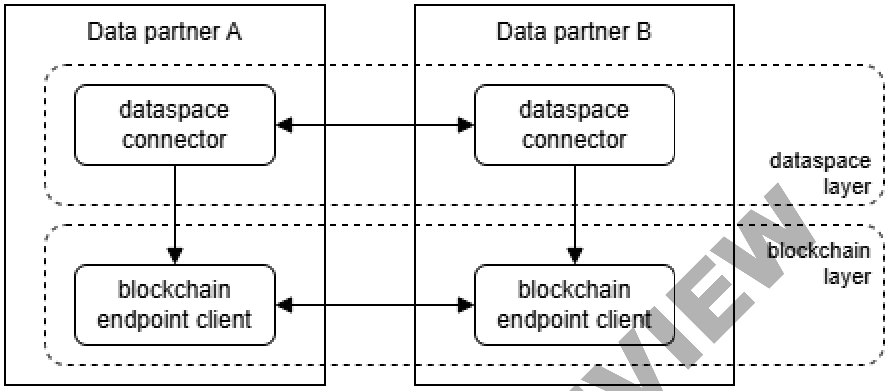

Figure 8: Overview of blockchain layer's conceptual architecture

4.2.1 BLOCKCHAIN ENDPOINT CLIENT

In terms of functionality, the blockchain endpoint client will support the following operations: Provide a smart contracts coupler to connect to the dataspace connector, provide a blockchain node to allow the system to participate in the blockchain network, and provide a command-line interface to allow queries to the blockchain ledger. Each of these functionalities is contained in its own sub-component, to allow for modularity and extensibility according to future needs, and to isolate control to individual controllers as needed. A schematic of the blockchain endpoint client architecture is presented in Figure 9.

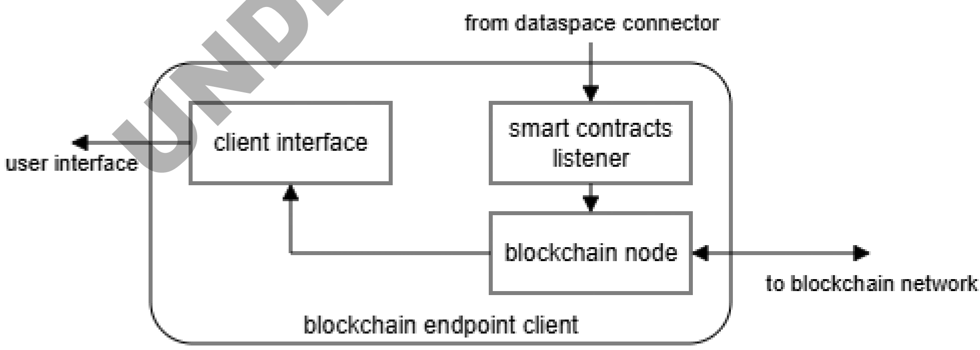

Figure 9: Blockchain endpoint client functional components

The smart contracts listener is coupled to the dataspace connector, and acts as the connection point to the dataspace. It awaits for events to be fired from the dataspace connector, and forwards these events to the blockchain node. The blockchain node performs multiple roles: first, it receives calls from the smart contracts listener, and forwards any payload received (that contains dataspace operations meta-data) to the rest of the blockchain network, in order to be added to the ledger, verified, and shared with the rest of the blockchain nodes (and thus also back to itself). And second, it awaits for queries from the client interface, runs those queries on the blockchain ledger, and returns all recorded transactions that fulfil the query criteria. Finally, the client interface provides a

Funded by the European Union

Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

simplified interface for external users (i.e. without requiring deep technical knowledge of the blockchain) to perform queries on the blockchain.

4.2.2 CONNECTION TO THE DATASPACE

As mentioned previously, the connection between the blockchain and the dataspace is realized via the blockchain endpoint client, and more specifically via the smart contracts listener. This module exposes an HTTP endpoint that expects a dataspace event to be passed as parameters, which it will then forward to the blockchain node module. In other words, the blockchain layer provides the infrastructure for dataspace events to be maintained in the blockchain, but it is the dataspace's responsibility to make use of this infrastructure.

4.3 NEXT STEPS

The architecture and functionalities described above have been developed and tested in a proof-of-concept environment, incorporating the minimum functionalities that demonstrate the viability of the system and establishing that a connection between a Data Space and a blockchain network is feasible. In order to do so, some constraints needed to be implemented, when compared to the planned final version. More specifically, at this stage the smart contracts blockchain endpoint client is not connected to the Data Space connector described in Chapter 5, but instead we have used a sample EDC deployment, for testing purposes. Additionally, both the sample EDC and the blockchain network have been deployed locally at this stage. And finally the client interface part of the blockchain endpoint client that is meant to support queries on the blockchain currently only implements a single query, to return all transactions currently on the blockchain ledger.

Moving towards the final planned version, the work plan for the coming months will address all of the limitations described here. More specifically, the blockchain endpoint client will be adapted to work seamlessly with the Dataspace connector that is being developed as part of Task 3.6. Furthermore, the blockchain network that will support the smart contracts functionality will be deployed on multiple machines over a remote network, corresponding to partners participating in the Data Space. Finally, for the user interface the client interface will expose an extensive API that will support performing detailed queries on the transaction history stored in the blockchain, including filtering criteria. The final format and supported queries of the API will be determined at a later stage, once the detailed use case requirements have been identified.

Funded by the European Union

Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

5 BIG-DATA ENABLED DATA SPACE IMPLEMENTATION

Data spaces are built on data sovereignty, interoperability, and controlled data sharing, with so-called "connectors" serving as the core mechanism for secure, policy-compliant data exchange among participants. This section aims to describe the dataspace connectors and their deployment in FEDORA, addressing multi-partner connectivity, and other important questions such as the integration with the public catalogue and Big-data support.

The approach aligns with the EMDS by supporting interoperable, federated data sharing between multiple mobility stakeholders. The dataspace connectors enable secure and standardized connectivity across participants. Moreover, integration with a central catalogue supports discoverability and alignment with common metadata and interoperability principles such as mobilityDCAT-AP, following the recommendations of the EMDS and the deployEMDS project for facilitating data discovery on a European level[4].

5.1 THE DATA SPACE CONNECTOR

5.1.1 DESCRIPTION

The connector is the core technical component that enables organisations to participate in a Data Space. It is a small software component installed on a participant's own systems that makes it possible to interact with other members of the Data Space. The dataspace connector is typically structured according to a separation between **a control plane** and **a data plane**:

- the control plane is responsible for managing the negotiation, governance, and orchestration aspects of data exchange, including participant authentication, policy management, contract negotiation, and discovery of available resources through catalogues,
- the data plane handles the actual transfer of data between participants, ensuring secure, policy-compliant, and direct exchange of datasets or streams between provider and consumer systems.

This separation allows governance to be managed independently from data movement and ensures that data remains under the control of its owner and is only shared under explicitly agreed conditions.

The Dataspace connector enables organisations to join the FEDORA Data Space and perform all key operations related to data sharing, handles the secure transfer of data assets between participants. It logs contracts and transactions for transparency and traceability, and manages the exchange of information about data offers, including terms and conditions of use.

5.1.2 LINK WITH THE MINIMUM VIABLE DATASPACE

The MVD provides the initial operational framework required to validate dataspace concepts, governance mechanisms, and interoperability principles among a limited set of participants. Within this framework, the dataspace connector acts as the technical enabler that allows organizations to securely exchange data while enforcing agreed-upon policies and trust mechanisms. In the

[4] https://deployemds.eu/wp-content/uploads/2025/07/D2.3-Technical-specifications-report.pdf

context of the MVD, dataspace connectors are deployed by participating organizations to establish trusted communication channels, publish and discover available data resources, negotiate data access agreements, and execute data transfers.

More specifically:

- Each partner deploys a Dataspace Connector,
- Each partner registers their datasets (e.g., traffic counts, traffic flows, OD flows), their usage policies (e.g., “only aggregated use”, “no resale”).
- The connectors negotiate data exchange contracts
- Data is exchanged directly between connectors
- Governance is enforced at runtime

5.2 MVD DEPLOYMENT

5.2.1 INITIAL DEPLOYMENT AT THE DATASPACE OPERATOR

In FEDORA, Akkodis as WP3 leader acts as the Dataspace operator responsible for managing participant onboarding, operating shared services, hosting the catalogues, hosting the issuer service, etc. After elicitation with the technical partners of the technological stack for the dataspace implementation, namely Eclipse Dataspace Components (EDC, more details in section 6.3), Akkodis deployed an EDC-based Minimum Viable Dataspace in a dedicated Azure cloud subscription, accessible at dataspace-fedora-horizon.eu.

5.2.2 CONNECTION GUIDE FOR PARTNERS

The FEDORA Data Space operator (Akkodis) then issued a procedure, describing how any FEDORA consortium partner could connect its own dataspace connector to this initial FEDORA connector (see Annex). This guide[5] is intended for technical teams at partner organizations that want to consume or exchange data via the FEDORA Data Space.

5.2.2.1 PREREQUISITES

Each partner intending to join the dataspace must meet the following prerequisites:

- Deploy an EDC connector on-premises using a supported orchestration platform, such as Kubernetes, Nomad, Docker Compose, or a bare-metal environment.
- Use DSP and DCP versions that are compatible with the MVD reference implementation.
- Provide a publicly resolvable DID, typically in the form of did:web:<their-domain>, with the corresponding DID document hosted and served over HTTPS at /.well-known/did.json.
- A valid MembershipCredential issued by the dataspace authority through the FEDORA issuer service (hosted by Akkodis).

The partner faces no restrictions regarding:

[5] Accessible here : https://dataspace-fedora-horizon.eu/partners

Funded by the European Union

Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

- The orchestrator or the cloud (EDC is interoperable at the protocol level, i.e. DSP + DCP via HTTP/JSON-LD: Nomad and Kubernetes are transparent to each other)
- The connector's internal database.
- The exact version of EDC (as long as DSP and DCP are compatible).

5.2.2.2 ENROLLMENT PROCEDURE

The partner provides Akkodis with:

- Their full DID (e.g., did:web:partner.example.com)
- The public URL of their DID document (verify that it responds via HTTPS without any certificate errors)
- A description of their role in the consortium (consumer, provider, or both)
- A technical contact

Akkodis as FEDORA's dataspace authority issues a MembershipCredential. The issuing service signs a Verifiable Credential (VC) of the following type:

```
{
  "@context": ["https://www.w3.org/ns/credentials/v2"],
  "type": ["VerifiableCredential", "MembershipCredential"],
  "issuer": "did:web:fedora-edc.francecentral.cloudapp.azure.com",
  "credentialSubject": {
    "id": "did:web:partner.example.com",
    "membershipType": "FedoraConsortiumMember",
    "validFrom": "2026-XX-XX"
  }
}
```

The signed VC is transmitted to the partner (typically via a secure retrieval endpoint provided by the issuer service). The partner stores the VC in its identity hub. On the partner side, the VC is ingested into the local identity hub.

For each outbound request to FEDORA, the partner connector performs the following steps:

1. Creates a VerifiablePresentation (VP) that includes this VC.
2. Signs the VP using its private key associated with the corresponding DID
3. Attaches the signed VP to the DSP request

5.2.2.3 VERIFICATION ON THE OPERATOR SIDE

The FEDORA connector performs the following validation steps:

1. Resolves the partner's DID to retrieve the associated public key from the DID document served over HTTPS
2. Verifies the signature of the Verifiable Presentation (VP).
3. Validates that the MembershipCredential VC has been correctly issued and signed by the issuer-service.
4. Enforces the EDC policy associated with the requested resource.

Once enrolled, the partner can request the connector catalogue:

```
curl -X POST https://<target-connector>/api/management/v3/catalog/request \
  -H "Content-Type: application/json" \
  -H "X-Api-Key: <token-mgmt-partner>" \
  -d '{
    "@context": { "edc": "https://w3id.org/edc/v0.0.1/ns/" },
    "@type": "CatalogRequest",
    "counterPartyAddress": "https://fedora-edc.francecentral.cloudapp.azure.com/api/dsp",
```

```
    "counterPartyId": "did:web:fedora-edc.francecentral.cloudapp.azure.com",
    "protocol": "dataspace-protocol-http"
  }'
```

5.2.2.4 FULL FLOW: FROM EDC CATALOGUE TO DATA TRANSFER

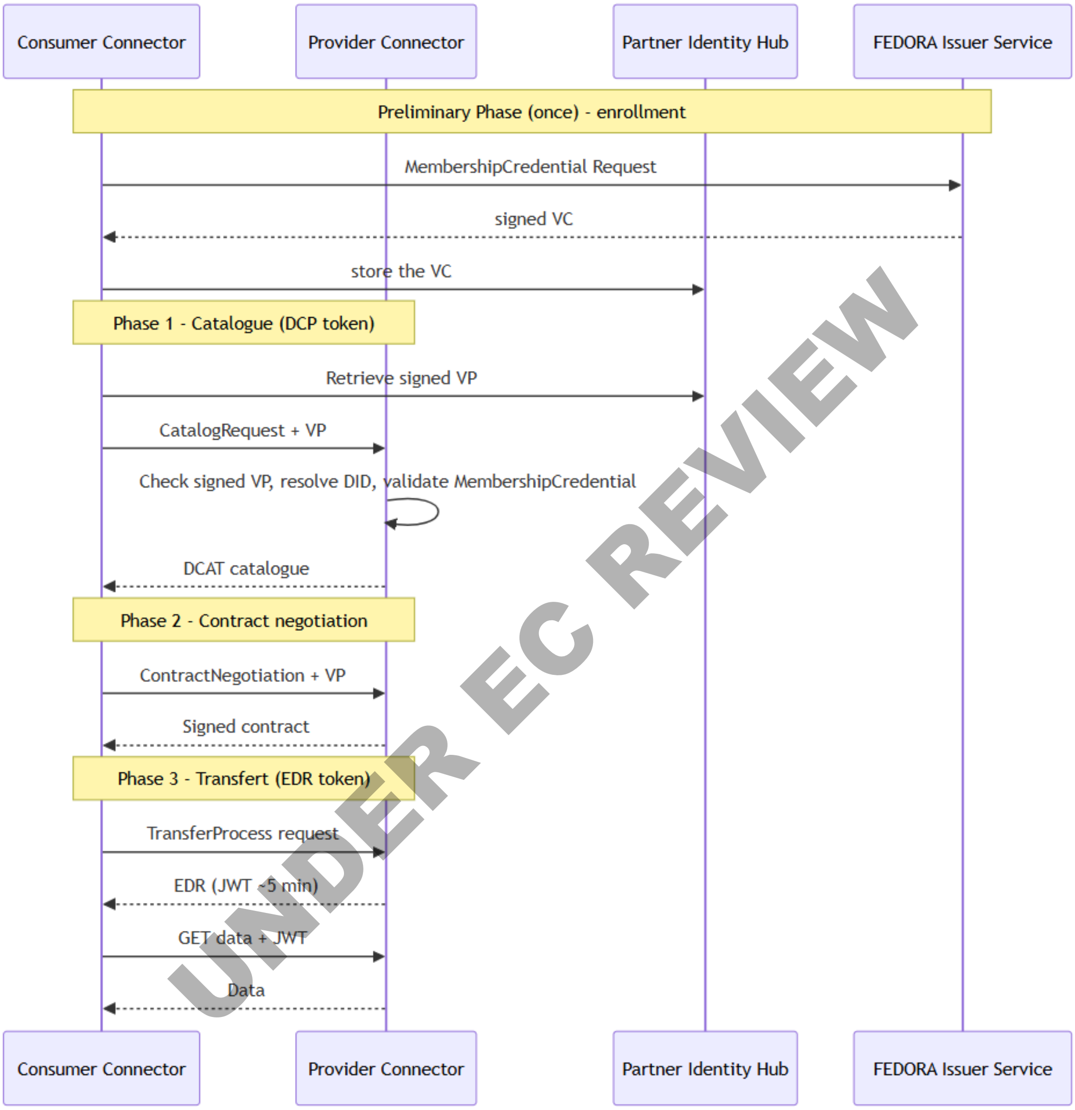

Figure 10: Sequence Diagram

5.2.3 FROM HANDS-ON SESSION TO A CORE FEDORA MVD

Following the deployment of the connector as standalone component within Akkodis for initial testing purposes, a group of three WP3 partners (namely Akkodis, ITS-Vienna, Frontier Innovations) then applied to form the “core FEDORA MVD”. These three core MVD partners progressively established connectivity between their respective environments to validate the fundamental mechanisms required for the dataspace interoperability, including basic data exchange workflows and connectivity patterns. This

Funded by the European Union

Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

took the form of a working session aiming to configure collaboratively the three respective EDC instances in a multi-partner environment, namely the “Hands-on EDC configuration”.

During this session, focused on connecting members’ systems through dataspace connectors, an open-access configuration was tested to validate interoperability and data exchange mechanisms. At this stage, the objective was to establish a working ecosystem with minimal complexity while enabling participants to publish and consume data under common rules.

Several iterations were necessary, showing the complexity of setting up a multi-partner connected environment, where alignment between stakeholders and technical components cannot be achieved in a single step. In fact, even though the term "minimum viable" of the MVD suggests something lightweight, the "minimum" part is about scope, not complexity reduction.

In a following iteration, additional trust services will be introduced, i.e. an issuer service for verifiable credentials operated by Akkodis, allowing the participants and their assets to be identified and authenticated through decentralized trust mechanisms.

5.3 INTEGRATION WITH THE PUBLIC CATALOGUE

As outlined in the previous paragraphs, dataspace connectors allow providers to publish metadata and offerings, and consumers can search and discover those offerings via their own connector. On the other hand, proven data catalogues such as CKAN also aims at metadata publication, playing a crucial role in enabling data discovery and access through data portals.

FEDORA aims to combine these two levels of cataloguing, resulting in an integration gateway capable of connecting to both open data portals and restricted data offers exposed via dataspace connectors. This dual capability makes the FEDORA Data Space a hybrid data ecosystem, combining open and controlled data within a unified discovery interface.

The public catalogue (CKAN) continues to manage and expose metadata (often in DCAT-AP) collected automatically from relevant external data portals, while EDC introduces a layer for access control, contract negotiation, and transaction traceability, thereby enabling access to potentially sensitive or private data without duplication.

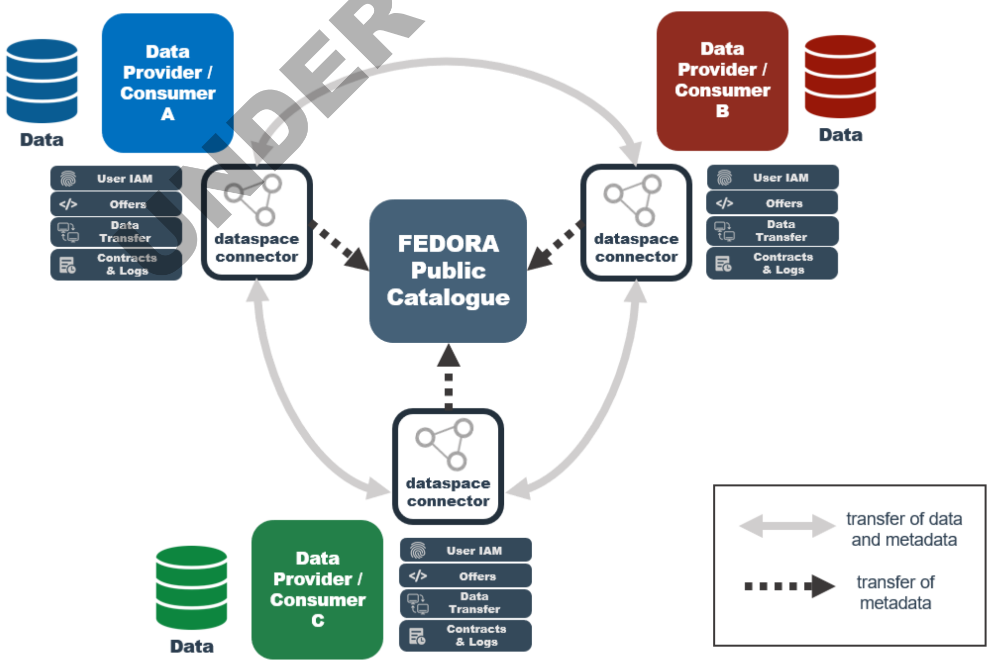

Figure 11: Data and metadata flows in the FEDORA dataspace

Funded by the European Union

Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

The same resource may appear in both catalogues, but with different levels of information and access rights: for example, a proprietary API may be documented in CKAN for discovery and listed in EDC only when it is integrated into a process involving sensitive data. This two-tiered approach reconciles public discovery with secure and controlled sharing, while ensuring metadata consistency and usage traceability.

This hybrid approach addresses the limitations of centralized architectures by enabling greater interoperability and by facilitating the implementation of “data sharing by design” models, which are particularly relevant in the mobility domain where multiple stakeholders must share data in a secure and governed manner.

5.4 BIG DATA SUPPORT

Big data is generally divided into two main categories: big static data and big stream data. Big static data refers to data that has already been stored, is relatively stable, and is analysed off-line. It is processed in batches and is primarily used to generate historical analyses or indicators. Conversely, big stream data refers to data produced continuously, arriving in streams and requiring processing in real time or near real time. Examples include data from sensors or GPS locations (e.g. of vehicles and/or app users). The fundamental difference is therefore temporal and operational: big static data allows to analyse what has already happened, while big stream data is used to react to what is currently happening.

Our work here focused on big static data sharing rather than streaming data sharing because static data offered a more practical and representative setting for exploring EDC-based data exchange. Static datasets are easier to process, allowing us to concentrate on the core dataspace mechanisms — such as participant interaction, policy enforcement, access control, and secure transfer — without introducing the additional complexity of real-time streaming infrastructure. Moreover, the data sharing use cases identified so far in FEDORA include historical sensor data and geo-referenced data, which involve the exchange of large but relatively stable datasets.

5.4.1 BIG STATIC DATA

In this first version of the FEDORA Data Space, we explored EDC-based large-scale static data sharing with a specific data source, namely a ClickHouse[6] database containing GPS trajectory data, and we focused on the practical challenges of efficiently sharing significant volumes of historical geospatial data across dataspace participants. ClickHouse is an OLAP database specifically designed to handle large amounts of data, using “highly optimized compression techniques and vectorized query execution to maximize CPU efficiency”.

Our approach was to create a filtered or anonymized view in the database (containing trajectory data), restricting columns, spatial scope, and level of detail. This view became the actual data asset to be shared and was then exposed through an intermediate API instead of direct SQL access. The API was registered in EDC as an asset, making it discoverable and accessible within the dataspace.

This layer enables additional processing such as anonymization, which will be explored in the next steps (see below).

[6] ClickHouse, https://clickhouse.com/

Funded by the European Union
Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

5.4.2 BIG STREAM DATA

Dataspace components are typically not designed as streaming engines such as Apache Kafka themselves. Instead, they focus on enabling sovereign data sharing through policy enforcement, identity management, and contract-based access control, while delegating the actual data transfer to dedicated components.

Moreover, directly exposing a Kafka broker to external participants via the connector would not be aligned with dataspace principles, because the connector would have to transform these streams before sharing them (e.g. by applying filtering, aggregation, or anonymization). In practice, transforming continuous streams before sharing would introduce significant architectural complexity and latency. We have not retained this approach at this stage.

5.5 NEXT STEPS

5.5.1 VERIFIABLE CREDENTIALS AND POLICY-BASED ACCESS CONTROL

The next phase will consist in establishing and implementing the decentralised trust and access-control framework for the FEDORA dataspace. This will include deploying issuing services to provide participating organisations with Verifiable Credentials, defining the mechanisms for organisational identification and authentication, and integrating these trust mechanisms with policy-based access control. The work will focus on enabling trusted organisations to access data, services, and other digital assets based on their verified credentials, attributes, and applicable access policies.

5.5.2 INTEGRATION OF PRIVACY-PRESERVING DATA ENRICHMENT

According to D3.1, the Mobility Data Harmonisation and Enrichment for Standard Exchange (MDHESE), in its local setup, will enable users to execute anonymization methods directly within their own infrastructure, ensuring that “sensitive data is fully anonymized before any cloud transmission takes place”. The next step will involve an integrated dataspace approach, in which data providers will be able to define different data products—corresponding to different levels of anonymization and/or aggregation with the MDHESE—and make them available to different data consumers, based on the needs identified in the pilot sites.

5.5.3 INTERCONNECTION WITH EXISTING DATA SPACES

In the next phase, interconnection with already existing (mobility) data spaces will be tested, leveraging the EMDS interlinking layer where possible, in order to facilitate interoperable data discovery, and cross-dataspace data sharing.

Funded by the European Union

Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

6 SOFTWARE STACK (TECHNOLOGICAL VIEWPOINT)

6.1 METADATA CATALOGUE SOLUTION (CKAN)

CKAN (Comprehensive Knowledge Archive Network) is the public-facing metadata catalogue to discover datasets in the data space in a user-friendly way. It is an open solution used by many transport and European national authorities as their official open data portal. CKAN is a compatible with the core requirements of the FEDORA data space catalogue which includes metadata discovery, standardized metadata display and access, harvesting, and access control. This catalogue was chosen among other metadata catalogues due to its modularity, the number of available extensions, and the variety of its plugins, but also based on the experience on the results of the project MobiDataLab (please refer to the D4.3[7] (V1) & D4.4[8] Reference Data Catalogue (V2)). For instance, the DCAT and particularly DCAT-AP specifications which allow to harvest datasets from external sources.

6.1.1 POSITIONING OF CKAN COMPARED TO PROPRIETARY SAAS PLATFORMS (SOCRATA, OPENDATASOFT) AND OPEN-SOURCE COMPETITORS (GEONETWORK, DKAN)

From the technological perspective of the RM-ODP framework, the choice of the CKAN solution is based on a comparative analysis of software components considering the requirements for distribution, interoperability, and managing the complexity of open systems. From a technological perspective, which aims to specify choices regarding infrastructure, platforms, and specific components, CKAN offers favourable characteristics as a foundation for a metadata catalogue. Its maturity, widespread adoption, and plugin ecosystem enable to effectively address deployment constraints in distributed environments, while offering native capabilities for metadata management, indexing, and exposure via REST APIs.

Compared to the Huwise (former OpenDataSoft) and Socrata platforms, CKAN positions itself as a fully customizable open-source framework, enabling deployment across various infrastructures (on-premises, cloud, hybrid) and fine-tuning to meet the specific requirements of the target system. In contrast, Socrata follows an “integrated platform” approach primarily offered as a SaaS solution, with no option for standalone deployment, which results in a strong dependency on the vendor for the system’s evolution, operation, and integration. From an RM-ODP perspective, this limits the ability to control the distribution, communication, and localization mechanisms of software components—elements that are nonetheless central to the engineering view. As for Huwise, the platform also follows a predominantly SaaS model, with a strong focus on ease of data publication, visualization, and utilization. Although it offers polished user experience and advanced distribution features, it has limitations similar to Socrata in terms of infrastructure control and deep customization, particularly in contexts requiring specific processing or complex distributed architecture. Its positioning is thus geared more toward “turnkey” use cases than toward building highly integrated modular systems.

Compared to other open-source solutions such as GeoNetwork or Drupal-based Knowledge Archive Network (DKAN), CKAN stands out for its more flexible, extension-oriented architecture, which facilitates the integration of specific components such as advanced harvesting, transformation, or dynamic translation mechanisms. GeoNetwork, while particularly effective for managing geospatial

[7] https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5ef7069c1&appId=PPGMS
[8] https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5052558ac&appId=PPGMS

Funded by the European Union
Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

metadata and compliant with ISO standards, has a specialized focus that can limit its extensibility in multi-domain contexts or those requiring custom processing pipelines. DKAN, which is based on a Content Management System, introduces constraints related to its underlying technology and the management of scalability in highly distributed environments.

From an RM-ODP perspective, CKAN enables the separation of concerns between different views (business, information, computational, and engineering) through its modular architecture and well-defined interfaces. It facilitates the implementation of the engineering view (communications management, component distribution, fault tolerance) and the computational view (decomposition into interoperable services) without imposing strict constraints on underlying infrastructure choices. Thus, the choice of CKAN is justified by its ability to serve as a central technological component in a distributed data management system, while offering a relevant balance between generality, extensibility, and operational robustness. Another reason for choosing CKAN is its ability to minimize vendor lock-in and to integrate seamlessly into scalable distributed architectures, unlike proprietary SaaS solutions that prioritize ease of use at the expense of system control and modularity.

6.1.2 CKAN ARCHITECTURE

CKAN architecture separates responsibilities across presentation, application, data management, search, storage, and deployment layers. It follows a multi-tier web application architecture allowing maintainability, scalability, and extensibility by separating matters across different functional levels.

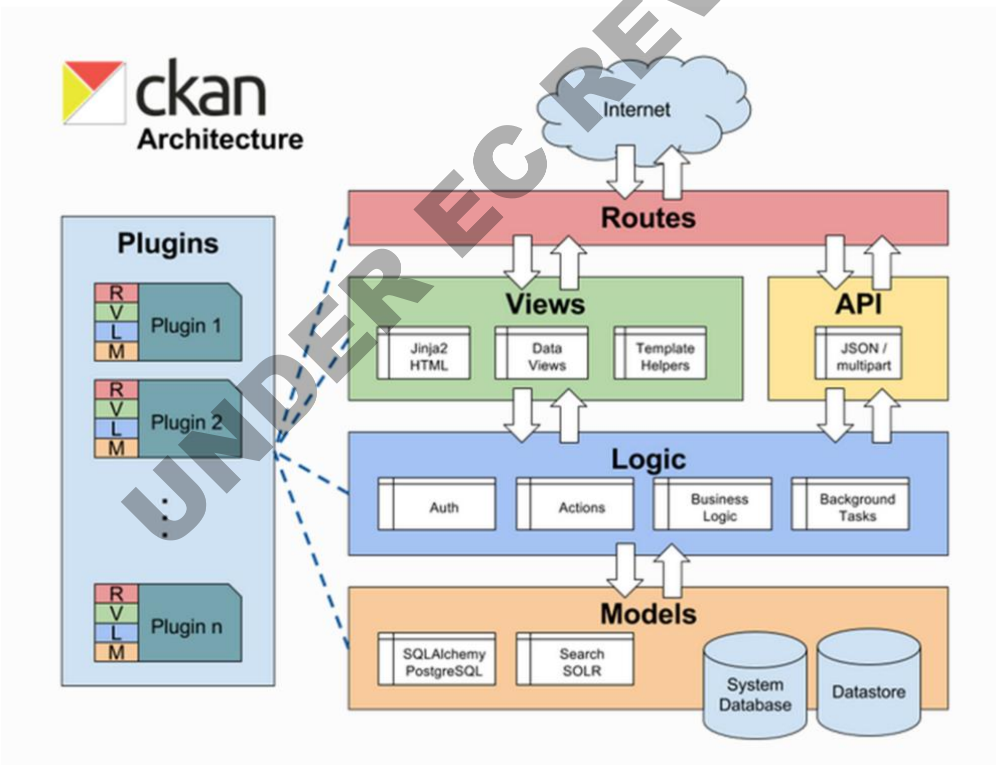

Figure 12: CKAN architecture diagram[9]

[9] source: https://docs.ckan.org/en/2.11/contributing/architecture.html

The presentation layer, also known as the Front-End, is responsible for user interaction and the visual rendering of the web interface. It ensures that users can browse datasets, perform searches, upload resources, and manage accounts through an intuitive graphical interface. The presentation layer of CKAN is independent from the application layer which promotes maintainability.

The application layer, also known as the Back-End, contains the core logic of CKAN and processes all user requests. It handles authentication, authorization, dataset management, API services, and plugin integration, and is developed in Python. A defining feature of CKAN's application layer is its plugin architecture. This one allows to extend CKAN's functionalities by creating plugins that modify templates, add new API endpoints, or introduce additional authentication methods. This modular design ensures extensibility without altering the core codebase, thereby supporting long-term sustainability and community-driven innovation.

The data access layer mediates communication between the application logic and the underlying database systems. It ensures structured storage, retrieval, and integrity of metadata and user information. This layer consists of PostgreSQL, Redis and Solr. The PostgreSQL database stores metadata records, user accounts, organizational information, access permissions, and configuration settings. It can be deployed via packages or containers. CKAN is integrated with Apache Solr, which is used as its search engine for fast discovery and exploration of datasets. This efficient search functionality is essential for data portals that manage large volumes of datasets (it is optimized for read-heavy workloads).

CKAN includes an authorization system that controls who can create, edit, and delete datasets, groups, and organizations. It enforces visibility rules such that only datasets authorized for publication appear in search results, while private datasets can only appear in the search results of logged-in users or members with the necessary permissions to access them. Permissions are managed through a combination of roles at the system level, the organizational level, or the dataset level. Unauthenticated users have a read-only access to public metadata; System operators have a full system-level access while registered organizations, editors, members or dataset collaborators have a particular access regulated by their role and the level of permissions provided by system admins and approved organizations admins.

6.1.3 DEPLOYMENT IN THE FEDORA PLATFORM

For the catalogue deployment, a Docker container was installed together with Supervisor, a process control system simplifying the installation and supporting operational consistency. Docker provides isolation and packaging, while Supervisor as a client/server system allows its users to monitor and control. The application server executing the CKAN backend code and communicating with the database and search engine was installed with Certbot for certificate management, providing secure and trust public access by obtaining issuance certificates from *Let's Encrypt*[10] *client*. The renewal of the service has been automated, following a periodic execution with a minimal runtime footprint.

6.2 AI STACK (RAG, DIFY)

There were different technological options for implementing RAG in the context of the FEDORA data catalogue, each offering distinct trade-offs in terms of scalability, maintainability, and integration with existing systems. Two primary options have been selected for detailed consideration and explored technically, both resulting in a concrete implementation:

- The first approach was an in-house solution combining a custom front-end application built with TanStack technologies and ChromaDB as the vector database. This solution allows full control over the data ingestion, embedding, and retrieval processes, enabling close integration with the catalogue's metadata, use case annotations, and any domain-specific rules.

[10] https://letsencrypt.org/

Funded by the European Union

Project funded by Schweizerische Eidgenossenschaft Confédération suisse Confederazione Svizzera Confederaziun svizra Swiss Confederation Federal Department of Economic Affairs, Education and Research EAER State Secretariat for Education, Research and Innovation SERI

The in-house design also permits the customization of the query and ranking logic, the implementation of advanced embedding pipelines for specific mobility datasets, and direct adaptation to evolving user requirements or data schemas.

- The second approach involved leveraging Dify, a platform that provides integrated vector storage and retrieval capabilities. Dify offers a managed environment with built-in support for RAG pipelines, reducing the operational overhead associated with managing embeddings, vector search indices, and model inference infrastructure. By relying on Dify, we could focus on configuring retrieval strategies and prompt templates, while benefiting from the platform’s optimization and scaling mechanisms. This approach provides rapid deployment and lower initial maintenance effort but offers less flexibility in terms of customization and control over the retrieval and embedding processes compared to the in-house solution.

Both approaches share the same fundamental architecture of RAG, involving the generation of embeddings from dataset metadata and associated documentation, storage in a vector database, and retrieval of relevant information in response to user queries for dynamic text generation by a large language model. The in-house solution emphasizes control, adaptability, and deep integration with the catalogue’s custom workflows, while the Dify approach prioritizes speed of deployment, operational simplicity, and reliance on managed services. The selection of either option depends on the balance between fully customised functionality and operational efficiency, as well as long-term maintenance. At the time of writing this document, we have opted for the Dify approach to ensure robustness and fast deployment.

6.3 DATASPACE STACK (ECLIPSE DATASPACE COMPONENTS)

The implementation of the FEDORA Data Space relies on the Eclipse Dataspace Connector (EDC), which provides the most viable foundation due to its performance in core business capabilities and its alignment with decentralized identity standards compared to other dataspace stacks.

This comparison relies on the outcomes of the deployEMDS project (i.e. the technical specification report[11]) which conducted a comprehensive assessment of existing software stacks and determined their suitability for the technical infrastructure of a mobility dataspace at the European level. The evaluation process involved comparing functional needs derived from customer journeys against actual stack implementations, focusing on core and extended building blocks. More specifically, the compliance of candidate stacks was assessed against six primary requirements: alignment with reference specifications, support from software providers, deployability, evolutionary path toward value-added services, compatibility with EU digital frameworks, and adherence to mobility data standards. Two stacks—EDC and FIWARE—were admitted to the testing facility. Several other candidates were analyzed but not initially admitted due to several reasons (lack of support for Verifiable Credentials, missing deployability requirements, components not yet available, etc.): iSHARE, Gaia-X/XFSC, Simpl, and Pontus-X. The evaluation prioritized components satisfying the most critical requirements for EMDS pilot projects, primarily focused on the Participant Agent (connector) and identity management.

The deployEMDS project currently focuses on complementing the EDC stack with additional components to fill identified gaps in user interface and onboarding self-service. Simpl remains a candidate for future piloting, particularly if its architecture evolves to allow for better interlinking with decentralized frameworks

[11] deployEMDS Consortium. (2025). *D2.3 Technical specifications report.* https://deployemds.eu/wp-content/uploads/2025/07/D2.3-Technical-specifications-report.pdf

Funded by the European Union

Project funded by Schweizerische Eidgenossenschaft Confédération suisse Confederazione Svizzera Confederaziun svizra Swiss Confederation Federal Department of Economic Affairs, Education and Research EAER State Secretariat for Education, Research and Innovation SERI

6.4 BLOCKCHAIN STACK (HYPERLEDGER FABRIC)

For the technical implementation of the blockchain layer of the FEDORA dataspace we propose the use of Hyperledger Fabric, an open-source distributed ledger technology platform. It covers all of the requirements outlined in Section 4, including: the execution of smart contracts (called “chaincode” in Fabric), peer nodes to be deployed on client systems and enable the “blockchain node” module described in Section 4.2.1, orderer nodes that ensure tamer-proof operation on the ledger, and consensus algorithms that allow all peers to verify the integrity of the ledger records. Given its capabilities for supporting the main requirements along with its open-source nature, we elected to develop the core blockchain functionalities of the FEDORA smart contracts module using Hyperledger Fabric.

In addition to Hyperledger Fabric acting as the core of the blockchain implementation, we make use of additional auxiliary technologies to enable the development and deployment of the smart contracts module. First, all sub-components of the blockchain endpoint client are developed on the Java language, one of the most widely used programming languages, which ensures compatibility with virtually all systems. This includes both the client interface and smart contracts listener, as well as any connecting logic within the blockchain endpoint client. Second, the full blockchain endpoint client module is wrapped within a Docker container, to ensure that deployment to partner systems is straightforward.

In terms of deployment to external systems, the smart contracts module is thus available as follows: A docker container is made available to download and run using the docker engine, which contains all of the logic to deploy an operational instance of the blockchain endpoint client. Additionally, a file containing all of the required environment variables will be provided, along with documentation on how to set up the necessary environment variables to connect the smart contracts module both to the dataspace connector and the rest of the blockchain network.

Funded by the European Union

Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

7 CONCLUSION

This deliverable has presented the design and implementation of the demonstrators developed in the first phase of the FEDORA Data Space: the Public Data Catalogue enhanced with an AI assistant, the Smart Contracts Layer, and the Minimum Viable Data Space. The report has highlighted their respective functionalities and their contribution to improved data discovery, transparent management of data-related transactions, and enabling trusted data sharing.

Overall, these demonstrators validate FEDORA's technical approach by demonstrating the feasibility of integrating interoperable data space technologies, intelligent data discovery services, and trusted transaction management into a unified platform. The results outline the project's contribution towards improving the accessibility, interoperability, and reuse of multimodal mobility data while supporting trustworthy collaboration among stakeholders.

In the next phase, we will further extend the adoption of the FEDORA Data Space among all relevant stakeholders, including possibly beyond the consortium. In doing so, we will explore a more bottom-up approach, complementing the mainly top-down approach followed so far, which has been driven by alignment with reference architectures and interoperability frameworks such as the EMDS. This complementary perspective will support the alignment with specific data needs and practical use cases from the pilot sites, while ensuring continued alignment with established data space principles and standards.

Funded by the European Union

Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

8 REFERENCES

Eclipse Foundation. (2026). Eclipse Dataspace Components. GitHub. https://github.com/eclipse-edc

Eclipse Foundation. (2026). Minimum Viable Dataspace [Source code]. GitHub. https://github.com/eclipse-edc/MinimumViableDataspace.

Eclipse Dataspace Working Group. (2025). Dataspace Protocol 2025-1. Eclipse Foundation AISBL. https://eclipse-dataspace-protocol-base.github.io/DataspaceProtocol/2025-1-err1/.

Eclipse Dataspace Working Group. (2025). Eclipse Decentralized Claims Protocol v1.0.1. Eclipse Foundation Specification Project. https://eclipse-dataspace-dcp.github.io/decentralized-claims-protocol/v1.0.1/.

W3C Credentials Community Group. (2026). did:web Method Specification. World Wide Web Consortium Community Group Draft. https://w3c-ccg.github.io/did-method-web/.

W3C Verifiable Credentials Working Group. (2025). Verifiable Credentials Data Model v2.0. World Wide Web Consortium (W3C) Recommendation. https://www.w3.org/TR/vc-data-model-2.0/.

Hyperledger Foundation. (2026). Hyperledger Fabric. Linux Foundation. https://www.lfdecentralizedtrust.org/projects/fabric.

Hyperledger Foundation. (2026). Hyperledger Fabric [Source code]. GitHub. https://github.com/hyperledger/fabric.

Funded by the European Union

Project funded by Schweizerische Eidgenossenschaft Confédération suisse Confederazione Svizzera Confederaziun svizra Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER State Secretariat for Education, Research and Innovation SERI

ANNEX - ONLINE SOFTWARE RESOURCES AND HOW-TO MANUALS

This annex provides references to the software artefacts developed within the FEDORA Data Space, including source code repositories and user documentation. The listed references include the location of the source code repositories, associated documentation (README files, user manuals, installation guides, and administrator guides), and, where applicable, the corresponding software release, version tag, or commit identifier. The referenced artefacts represent the versions available at the time of submission of this deliverable.

ANNEX A – SOFTWARE REPOSITORY REFERENCES

IDENTIFIER	DESCRIPTION	REPOSITORY	VERSION / COMMIT
SW-01	Project source code	https://github.com/fedora-horizon	Release v1.0
SW-02	Data Space documentation (README)	https://fedora-horizon.github.io/docs/	Main branch (accessed version)

ANNEX B – USER DOCUMENTATION

DOCUMENT ID	TITLE	LOCATION	VERSION
UM-01	User Manual – **How to find and collect data in the public-facing catalogue**	https://fedora-horizon.github.io/docs/harvesting/introduction/	Release v1.0
UM-02	User Manual – **How to discover and use data with the AI assistant**	https://fedora-horizon.github.io/docs/ai-assistant/introduction/	Release v1.0
UM-03	User Manual – **How a consumer connector retrieves data from a provider connector over the EDC protocol.**	https://dataspace-fedora-horizon.eu/demo	Release v1.0

ANNEX C – DEVELOPER DOCUMENTATION

Funded by the European Union

Project funded by
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

DOCUMENT ID	TITLE	LOCATION	VERSION
DG-01	Partner connection guide for the FEDORA EDC connector	https://dataspace-fedora-horizon.eu/partners	Release v1.0

Funded by the European Union

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI