Politecnico di Torino

Politecnico di Torino

Master's Degree in Engineering and Management

A.Y. 2024/2025

Graduation Session November 2025

Blockchain Integration in Data Space Architectures: the NOUS Project

Ensuring Integrity, Notarization, and Accountability

Supervisors:

Prof. Guido Perboli
Prof. Marco Gajetti
R.A. Fabiano Columbano

External Supervisors:

Prof. Marco Mamei

Candidate:

Federica Gentili

Abstract

The exponential growth of data and the emergence of federated digital ecosystems have made data spaces a cornerstone of the European Data Strategy. While these environments enable interoperability and sovereignty, they still rely on implicit trust among autonomous participants. This thesis investigates how blockchain technology can complement governance in data spaces by introducing verifiable mechanisms for integrity, traceability, and accountability. Within the context of the EU NOUS Project, the thesis proposes a blockchain-enabled architecture for data lifecycle management in federated data spaces. The proposed Proof of Concept (PoC) integrates blockchain notarization into the operational workflow of a data space, ensuring that each operation is traceable and immutable. The approach is validated through a top-down methodology, moving from architectural design to practical implementation that combines the data space middleware, EBSI-based blockchain layer, and the MASA use case. The results demonstrate that integrating blockchain strengthens the transparency and trustworthiness of federated data ecosystems. This framework takes shape within Europe's ongoing advancement toward trusted and interoperable data sharing, contributing to the vision of sovereign, accountable, and verifiable digital infrastructures.

Acknowledgements

Here I wish to express my gratitude to all those who have, in different ways, contributed to this meaningful achievement, that opens the way to what comes next.

I am deeply thankful to my supervisor, Prof. Perboli, as well as to Prof. Gajetti, Prof. Mamei, and R.A. Columbano for granting me the opportunity to be part of this project, and for the guidance, insight, and support they have shared with me throughout this journey.

My heartfelt thanks go to my parents and to my entire family. I am grateful for your love and the steady source of strength and inspiration you have always been.

I also wish to thank my partner, for the road we continue to walk hand in hand and for reminding me of the profound beauty found in life's simplest things.

Finally, my warmest thanks go to my friends - those from childhood, those I met throughout my university years, and those encountered during my experiences abroad. Each of you has enriched me. I will hold close the memories, lessons, and kindness you have given me, with deep gratitude and affection.

Above all, I am grateful for the journey these years have been: for every encounter, every challenge, and every moment that has brought me to where I stand today.

Federica

Table of Contents

List of Figures v

1 Introduction 1
- 1.1 Motivation and Objectives . . . 2
- 1.2 Contributions . . . 2
- 1.3 Structure of the Thesis . . . 3

2 State of Art 5
- 2.1 Data Management in the Era of Dataspaces . . . 5
  - 2.1.1 From Centralized to Federated Data Infrastructure . . . 5
  - 2.1.2 Dataspaces . . . 6
  - 2.1.3 Core Principles of Dataspaces . . . 9
  - 2.1.4 Structure and Components of a Dataspace . . . 12
  - 2.1.5 Roles in a Dataspace . . . 15
  - 2.1.6 Ongoing Cloud-to-Edge and Dataspace Initiatives . . . 16
- 2.2 Blockchain Technology . . . 20
  - 2.2.1 Origins . . . 20
  - 2.2.2 Core Principles . . . 21
  - 2.2.3 Structural Elements . . . 23
  - 2.2.4 Cryptography . . . 25
  - 2.2.5 Consensus Algorithms . . . 27
  - 2.2.6 Smart Contracts . . . 29
  - 2.2.7 The Scalability Trilemma . . . 30
- 2.3 Edge Computing and Smart Cities . . . 31
  - 2.3.1 Smart Cities and the Evolution of Intelligent Mobility . . . 31
  - 2.3.2 Safety and the Protection of Vulnerable Road Users . . . 32
  - 2.3.3 Multi-Access Edge Computing: Concept and Capabilities . . . 33
  - 2.3.4 Edge Computing in Smart City Infrastructures . . . 34
  - 2.3.5 The Synergy of Edge Computing, 5G, and V2X Networks . . . 35
  - 2.3.6 Applications and Real-world Implementations . . . 36
  - 2.3.7 Challenges and Future Directions . . . 36

3 The NOUS Project 38
3.1 Cloud Computing and the European Data Strategy 38
3.2 Overview of NOUS project 40
3.2.1 NOUS Vision and Strategic Objectives 41
3.3 NOUS Architectural Design 42
3.4 Use Cases 44
3.4.1 MASA Use Case: Edge Computing for Connected and Safe Mobility 46

4 A Blockchain Framework for Data Lifecycle Management in Data Spaces 52
4.1 System Architecture 52
4.1.1 The European Blockchain Services Infrastructure - EBSI . . 56
4.1.2 Simpl-Open: A Federated Middleware Framework for European Data Interoperability 57
4.1.3 Processes Design 61
4.2 Proof of Concept (PoC) Implementation Steps 65
4.2.1 Blockchain Incremental Validation Framework 66
4.2.2 Evolution of the Interaction Layer 69
4.2.3 Data Provider Integration Layer 71
4.3 Current Implementation 73
4.3.1 Sovity Environment 73
4.3.2 API-Based Workflow 75
4.3.3 End-To-End Flow 77
4.4 Value Creation and Impact of the Proposed Framework 83

5 Conclusions and Future Work 87

A Appendix 90
A.1 GitHub Link 90
A.2 Glossary 90
A.3 Diagrams 93
A.4 PoC Current Implementation 97

Bibliography 99

List of Figures

2.1 Data First, Schema Later or Never Models 8
2.2 Dataspace Components [10] 12
2.3 Generic Chain of Blocks [26] 24

3.1 MASA's Smart Cameras [46] 46
3.2 MASA's Infrastructure Architectural Design [46] 47
3.3 MASA's Mobile Interface with Relevant Alerts and Experimental Evaluation 49
3.4 MASA's In-Vehicle Led Notifications [46] 49

4.1 System Architecture 53
4.2 Simpl-Open Architecture Vision: Actor Groups and Cross-Dataspace Interoperability 59
4.3 Blockchain Incremental Phases 67
4.4 Interaction Layer Incremental Phases 69
4.5 Webhook-Based Integration Workflow 77
4.6 Webhook Steps 78

A.1 UML Diagram: Onboarding Operation 93
A.2 UML Diagram: Add Operation 93
A.3 UML Diagram: Update Operation 94
A.4 UML Diagram: Withdrawal Operation 94
A.5 Sequence Diagram: Onboarding Operation 95
A.6 Sequence Diagram: Add Operation 95
A.7 Sequence Diagram: Update Operation 96
A.8 Sequence Diagram: Withdrawal Operation 96
A.9 Sovity Sandbox Dashboard 97
A.10 Webhook 97
A.11 Ganache Transaction 98

Chapter 1

Introduction

New paradigms for data sharing and governance have emerged in the last years due to the exponential growth of digital data and the rise of cross-domain data-driven services. Among them, data spaces are envisioned as federated ecosystems where participants (public and private organizations, institutions, and individuals) can exchange data under shared governance models while maintaining data sovereignty. The European Commission encourages the creation of interoperable and trustworthy data spaces that comply with key legislative frameworks, such as the General Data Protection Regulation (GDPR) and the Data Governance Act (DGA). Whereas the DGA sets the ground for reliable data intermediaries and procedures for data reuse, the GDPR ensures legal, fair, and transparent processing of personal data.Together, these frameworks set the ethical and legal foundation of the European data economy.

However, despite data spaces' potential, establishing trust and accountability among distributed and independent actors remains a key challenge. Due to this hurdle, there has been an increased emphasis on blockchain technologies, which provide decentralized, auditable, and tamper-resistant mechanism of recording and verifying transactions or events. Transparent ecosystems where each operation on data can be notarized and verified in a verifiable way can be created by merging the concepts of data spaces and blockchain.

Within the context of the European project NOUS, this thesis investigates the integration of blockchain technology and data spaces. In this regard, the current study explores how the blockchain technology can support the notarization and traceability of data operations carried out within a data space. The growing need for trust-enhancing mechanisms that can support current interoperability frameworks and offer unchangeable proof of data transactions without sacrificing effectiveness or privacy is what encourages the research.

This thesis addresses the need for verifiable trust among independent and federated actors by presenting a Proof of Concept (PoC) and reference architecture that emphasize the integration of blockchain for the notarization and traceability

of data operations.

1.1 Motivation and Objectives

This work is driven by the recognition that data spaces still heavily rely on participant trust assumptions, despite promoting interoperability through standardized interfaces and governance rules. Although policy-based trust frameworks govern data spaces, they are insufficient on their own to offer cryptographic assurances or externally auditable proof of data transactions. Participants are expected to comply with data usage policies, consent rules, and service-level agreements defined under governance contracts ; however, manual audits or central oversight can be required to confirm that these commitments are fulfilled.

The federated nature of data spaces, where each participant should retain autonomy without compromising collective trust, is incompatible with this dependency. By adding a decentralized trust layer where the crucial metadata of data operations, like publication, update, or access revocation, are recorded in a tamper-proof, verifiable ledger, blockchain technology can enhance current governance models. Rather than replacing data space governance, blockchain acts as a technological enforcer and validator of its principles, bridging the gap between policy trust and operational trust. The accountability mechanisms already established by frameworks like the Data Governance Act (DGA) and GDPR are strengthened by the fact that every notarized event becomes a piece of unchangeable evidence that can be independently verified by third parties, regulators, or auditors.

The following goals of the thesis are outlined based on these assumptions:

- Evaluate how blockchain features and data space principles align, identifying the advantages and limitations of integrating them in terms of scalability, compliance, and interoperability.
- Design and implement a Proof of Concept (PoC) that illustrates how blockchain technology can facilitate the traceability and notarization of data-related operations in a federated data space.
- Propose an incremental evolution of the PoC, aiming for advancements in the blockchain's underlying technology as well as the layers that connect the data space and blockchain.

1.2 Contributions

The contribution of this thesis is twofold. In order to show how blockchain can operationalize the principles of trust and transparency underpinning the European

data space model, it first provides a conceptual and architectural contribution. Second, it provides a technical artifact - a working Proof of Concept (PoC) - that acts as a concrete illustration of how data space and blockchain infrastructures can coexist and enhance one another.

The suggested solution enables external auditing and verifiability without central oversight, thus it strengthens accountability-by-design through connecting blockchain notarization to the operations identified within the data space middleware. Contributing to the broader debate on regulatory compliance and trust mechanisms under the DGA and GDPR, this work provides insights on how distributed ledgers can improve transparency in areas like smart mobility that are sensitive to compliance.

1.3 Structure of the Thesis

The thesis is organized into five main chapters, followed by appendices containing supporting materials and a glossary of technical terms. Each chapter contributes to building a coherent narrative that moves from the theoretical and contextual foundations of data spaces and blockchain technologies to the practical realization and evaluation of the proposed framework.

Chapter 2 - State of the Art: this chapter provides the conceptual and technological background necessary to frame the research. It explores the evolution from centralized to federated data infrastructures and defines the core principles, structure, and roles that characterize data spaces. It then presents the fundamental aspects of blockchain technology, including its architecture, cryptographic underpinnings, and consensus mechanisms, and concludes with an overview of edge computing and smart city paradigms. Together, these sections establish the multidisciplinary foundation upon which the proposed approach is built.

Chapter 3 - The NOUS Project: the third chapter introduces the NOUS Project, positioning it within the broader European data strategy and cloud-to-edge continuum. It describes the project's vision and strategic objectives, its architectural design, and its operational use cases. Particular attention is devoted to the MASA use case, which serves as the practical scenario through which the blockchain-enabled dataspace integration is developed and validated.

Chapter 4 - A Blockchain Framework for Data Lifecycle Management in Data Spaces: this chapter represents the core of the thesis and presents the architectural framework and Proof of Concept (PoC) implementation. A top-down approach is employed, starting from the overall system architecture and progressively detailing its components and operational layers. The chapter first introduces the architectural foundations (EBSI and Simpl-Open), then describes the design of the key processes

modeled within the dataspace context. It proceeds with the implementation aspects, including the incremental blockchain validation framework, the evolution of the interaction layer, and the integration of the data provider (MASA). The final sections focus on the current implementation, the API-based workflow reproducing the end-to-end execution of the system, and a concluding reflection on the value creation and impact of the proposed framework.

Chapter 5 - Conclusions and Future Work: the final chapter summarizes the main contributions and results achieved by the thesis and it traces future development directions.

Appendices: this section provides additional reference material, and a Glossary.

Chapter 2

State of Art

2.1 Data Management in the Era of Dataspaces

2.1.1 From Centralized to Federated Data Infrastructure

In the evolving landscape of digital economy, data has become a central resource, on the same level as physical capital and labor, in driving innovation, competitiveness and systemic transformation. Among the several categories of data utilization, inter-organizational data sharing has emerged as a key enabler of collaboration and value creation. Despite its strategic relevance, the notion of data sharing suffers from conceptual ambiguity and is often used interchangeably with related terms such as "data exchange", leading to inconsistent interpretations across domains. It is usually misused to represent other mechanisms such as data exchange or data transfer, which lack the structured governance, legal frameworks and interoperability requirements that real data sharing entails [1].

Data sharing involves the deliberate and governed provisioning of data to third parties - often across organizational boundaries - under defined conditions, policies and incentives [1]. It is typically supported by digital infrastructures and collaborative platforms that enable its trusted and purpose-driven reuse [2]. As opposed to data exchange, which is mainly the technical transformation and transfer of data between multiple systems, data sharing represents a broader, socio-technical process that entails contractual agreements, trust mechanisms and gives rise to shared infrastructures such as data marketplaces or platforms [1]. That is, data exchange is a technical subprocess of the broader data sharing, and focuses on how data is translated, moved and made interoperable between systems, while data sharing defines the overall practice and governance of making data available for reuse between entities [1]. This description is necessary to seize the multi-layered nature of modern data ecosystems, where successful sharing means not only the technical capacity to share data but also the institutional capacity to align rights,

responsibilities and incentives across stakeholders.

Historically, most digital infrastructures were built around centralized architectures, wherein all relevant data was ingested into a singular system (e.g., relational database) controlled by a central authority. These systems allowed for rigorous control and optimized performance in closed environments, but they required extensive semantic integration and up-front homogenization of data [3]. As organizations and datasets grew both in size and diversity, thus there is not a unique schema to which all data conforms, these centralized models became increasingly unfit to handle the need of modern and distributed data ecosystems [4]

To address this issue, the concept of *federated infrastructure* emerged. Rather than imposing uniform data schemas or central repositories, federated models allow interoperability and coordinated access while data remain within its original domain, managed by its rightful owner [3]. This *data co-existence* approach acknowledges the reality of today's data ecosystems, where information is often distributed across systems, formats, jurisdictions and owners [4]. Moreover, it significantly reduces the need for prior integration, enabling organizations to participate in data ecosystems without losing control over data but they agree to share through standardized protocols and shared governance principles. Federated models promote incremental integration, called pay-as-you-go model: models can be offered even when full semantic alignment is not yet achieved and data quality or availability varies across sources [4, 3].

A concrete manifestation of this paradigm shift is the concept of *dataspaces*. They refer to federated digital ecosystems where multiple participants collaborate around shared data assets under agreed rules and standards. Being now central to the European Strategy for Data, they aim to facilitate sovereign, interoperable and sector-specific data collaboration, supporting use cases that range from connected vehicles to industrial automation and scientific research [5, 1].

2.1.2 Dataspaces

The term *dataspaces* was introduced in 2005 by Maier et al. as a new abstraction for data management, implemented through Dataspace Support Platforms (DSSPs) [3]. These platforms support data sharing in multiple formats, accessibility across different systems and interoperability with diverse applications [6].

Dataspaces are "described simply as a set of relationships and participants" and "large-scale heterogeneous collection of data distributed in several sources in various formats, with a mechanism to handle structured, semi-structured, and unstructured data" [6]. Through a co-existence approach, dataspaces do not physically integrate the data; rather, the data remain stored at their source and are integrated at the semantic level. Moreover, nested dataspaces are frequent, allowing each participant to be part of several ecosystems [2].

Comparison among different data management approaches

In order to clearly define the notion of dataspaces, a comparison with the previous approaches is summarized in the following table [6].

Table 2.1: Comparison among different data management approaches

Name	**Description**	**Centralised / Decentralised / C / D**	**Structured Unstructured S / U**
Relational Database	Structured collection of data	C	S
Non-Relational Database	Organized collection of data	C+D	U
Data Warehouse	Central repositories of integrated data	C	S
Data Lake	Centralised repository to store and process large amounts of structured, semi-structured, and unstructured data	C	S+U
Spatial Data Infrastructure	Data infrastructure for spatial data utilisation	C+D	S
Data Space	Federation of decentralised data ecosystems	D	S+U
Data Mesh	Domain-oriented decentralised architecture	D	S+U

In the past decades, data management paradigms such as relational databases, data warehouses, and later data lakes, have provided structured ways for organizations to store, organize, and query data primarily for internal use. These systems were structured to ensure consistency, performance and centralized control [4].

However, as digital transformation advances, data ecosystems become increasingly interconnected, these traditional paradigms became inefficient to fully exploit the potential of data as an asset. According to Bacco et al. (2024),the unwritten law of *schema first, data later* typical of DBMSs indeed evolved into *data first, schema later or never*.

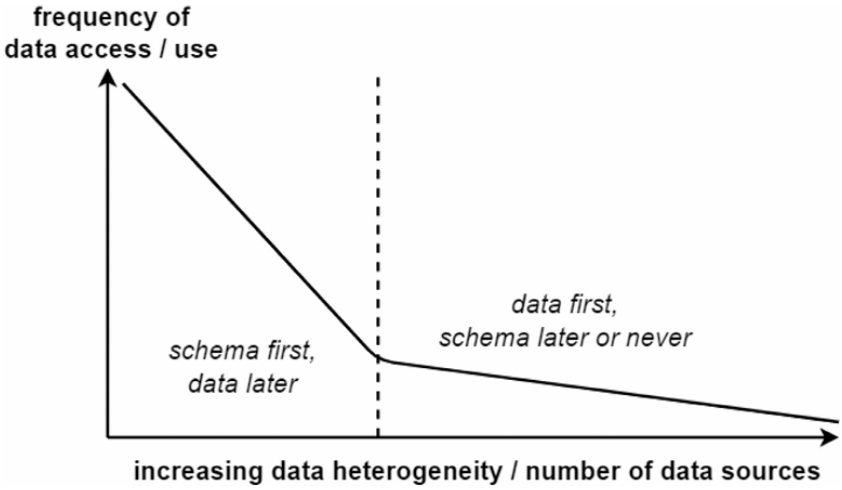

Figure 2.1: Data First, Schema Later or Never Models

Fig. 2.1 depicts the aforementioned shift in data management paradigms as a function of two variables: increased data heterogeneity and decreased access or reuse frequency. In situations where data is homogeneous, frequently accessed and strongly coupled with business processes, the *schema-first* paradigm - where the schema is established before data is gathered - is usually preferred. This applies to traditional relational database systems, which rely on strict adherence to pre-specified schemas for consistency, performance and reliability [6]. Where data becomes more heterogeneous in form, provenance and interpretation - and particularly where intermittent or cross-organizational use comes into play - this model is less efficient. Here a *data-first* logic comes to the fore, whereby data is made available before any hard structure is imposed or indeed at all. Frameworks such as data spaces are showing this shift: instead of integrating data at a single point of integration, they provide decentralized sharing through the application of metadata, access policies and semantic descriptions that can be dynamically interpreted. In this instance, the links among data elements are not necessarily

pre-established in a schema but can be inferred at usage time [6].

Moreover, the rise of the platform economy and the inclusion of industrial sectors in data-driven value chains challenge companies to consider data not only as an internal resource, but a strategic resource. Nowadays, data plays several roles in the economy: it enables operational excellence, it is commercialized as a product, it fosters innovation through cross-organizational collaboration and it is increasingly seen as a strategic resource for long-term sustainability [2]. Data only acquires value through use and reuse, which raises questions of data governance, trust and fair benefit sharing. These challenges demand new frameworks - such as data spaces - that allow for decentralized, sovereign and policy-driven data sharing across domains and actors.

2.1.3 Core Principles of Dataspaces

Having clarified the notion of dataspaces as a co-existence approach, we now outline the core principles that govern their technical and institutional design. The concept of data spaces, as defined within modern academic and policy literature, is based on a collection of core principles that distinguish them from previous data integration and data exchange practices. While these principles not only define the technical and governance structure of data spaces, they also provide the foundation for the creation of sovereign, interoperable, and trust-based data environments. Below are the most relevant principles underpinning current data space initiatives.

Data Sovereignty and Autonomy

Data sovereignty is a concept where data providers own and are in total charge of their data assets, including decisions around whom to authorize access to the data, for what, and under what conditions. Contrary to traditional data sharing arrangements or open data portals, where mastery over data is relinquished at some stage once data go public, data spaces enable participants to share data selectively, dynamically, and reversibly. Sovereignty in this context means both legal control and operational independence. That is, the ability to select with whom information is shared, limitations or conditions on use, and the capacity to revoke access at any moment in time. Autonomy, however, is not equivalent to irresponsibility. Participants must remain actively in control of the data-sharing process and capable of managing access decisions without undue reliance on intermediaries, except where required by legislation. The Data Spaces Manifesto highlights this double dynamic of agency and responsibility as being at the heart of trusted collaboration on data. This design principle is one of the pillars of European data initiatives and it is in immediate compliance with the European Strategy for Data [7, 8].

Decentralization, Neutrality and Federated Control

As a counterpoint to central data platforms, data spaces are structured as federated ecosystems. Each participant - whether a business firm, public agency, or research institution - is an independent node, supplying data and services and adhering to common rules and technical standards. This design avoids single points of control and enables models of governance based on equality and respect for each other. Decentralization is not only architectural, but political: no agent should need to override decisions or inhibit the agency of others. This neutrality ensures that participants engage on a level playing field, with the same rights and obligations in furnishing, requesting, and negotiating access to data. A data space's architecture is designed to be neutral, allowing for this balance[7, 8].

Interoperability

Interoperability is required to enable seamless data exchange and service integration between parties in a data space. It occurs on three levels: technical, semantic, and organizational. Technical interoperability is founded on standardized protocols, APIs, and data formats to enable machine-readable data exchange. Semantic interoperability ensures that common data is consistent in its meaning between contexts, typically through the use of ontologies, vocabularies, and metadata standards. Organizational interoperability is interested in aligning processes, responsibilities, and policies within institutions to facilitate successful collaboration. Interoperability is a requirement for data space reusability and scalability [7, 4].

Trust, Transparency and Accountability

Establishing trust within an open and distributed environment is central to the success of data spaces. Trust needs to be not only established by ethical intention but enforced and verifiable. Users must be able to rely on auditable trails, consent records, and carefully detailed use agreements. Distributed ledger technology and smart contracts can make this possible by enabling tamper-evident records and automated enforcement of rules. In line with the principle "act in good faith, but verify" stated in the Manifesto, trust must be supported by the technical capability of monitoring, authentication, and, where needed, reacting to breaks or violations. Every participant must be capable of ensuring other participants adhere to shared conditions and be empowered to suspend or end sharing in case of non-adherence. Such an accountability structure lead to a healthy and resilient data-sharing environment [7, 8].

FAIR Principles

The FAIR principles - Findability, Accessibility, Interoperability, and Reusability - evolved in scientific data management but are now central principles for data spaces in general. They promote data assets rich in metadata that are simple to find, accessible under defined terms, semantically interoperable, and organized for reuse over the long term. Implementing FAIR principles makes data searchable across disciplines and platforms while requiring data to be described and regulated in a way that facilitates its consistent use. European initiatives such as the European Open Science Cloud (EOSC) and national data infrastructures such as Germany's NFDI have adopted FAIR as an underpinning design principle [7, 9].

Private Channels and Infrastructure-Agnostic Sharing

The Manifesto highlights that data itself is shared through distinct peer-to-peer channels, outside the data space. The latter only sets the ground for policy definitions, agreements and metadata exchange between parties that follow the agreed protocols. As a consequence, the data space is platform- and technology-agnostic, facilitating scalability and inclusiveness of multiple technical configurations [8].

Flexibility of Business Model and Ecosystem

Data spaces are not business models in themselves or closed digital ecosystems. They are rather enabling infrastructures upon which various models and collaboration patterns can be formed. By providing the technological and contractual foundations for secure data exchange, they themselves constitute a middle layer, which enables innovation and adaptation. Commercial, scientific, or civic purposes, the prospects unleashed by data spaces are varied and not limited to any one set-up. This openness to support diverse models of value creation lies at the core of the long-term sustainability of the data space model [8].

Governance and Participation

Effective governance is needed to ensure that data spaces are inclusive, transparent and sustainable. Data space governance models tend to be predominantly a union of public and private players with clearly defined roles, responsibilities and deciding authorities. Such models need to ensure fair participation so that small and large players can contribute as well as benefit from it. Governance models also decide on rules for data access, monetization, conflict resolution, and technical conformity. Through aligning interests and strengthening accountability, robust governance mechanisms keep any single actor from controlling and preserve the federated

nature of the system [7, 6].

These guidelines collectively form the regulatory and operational foundation for data space design. They enable a balance between innovation and control, autonomy and cooperation, decentralization and interoperability. Data spaces are therefore a potential model for data governance in an increasingly interconnected world.

2.1.4 Structure and Components of a Dataspace

After an outline of the theoretical foundations of dataspaces, the discussion now turns to their structural characteristics, where the International Data Spaces Association (IDSA) has played an important role in defining their technical and governance foundation. The IDSA is a non-profit organization with industry, research and government representation that jointly develop standards and frameworks to enable sovereign, secure and interoperable data exchange. Its mission centers on building a trusted data economy by promoting architectures that respect data sovereignty but, simultaneously, are interoperable and scalable across sectors and domains [10]. The cornerstone of this effort is the IDS Reference Architecture Model (IDS-RAM), a comprehensive framework that formalizes the technical, operational and governance characteristics of data spaces. IDS-RAM outlines a decentralized, modular framework under which participants retain and keep control of their data and engage in trusted and secure exchanges through a federated ecosystem [11].

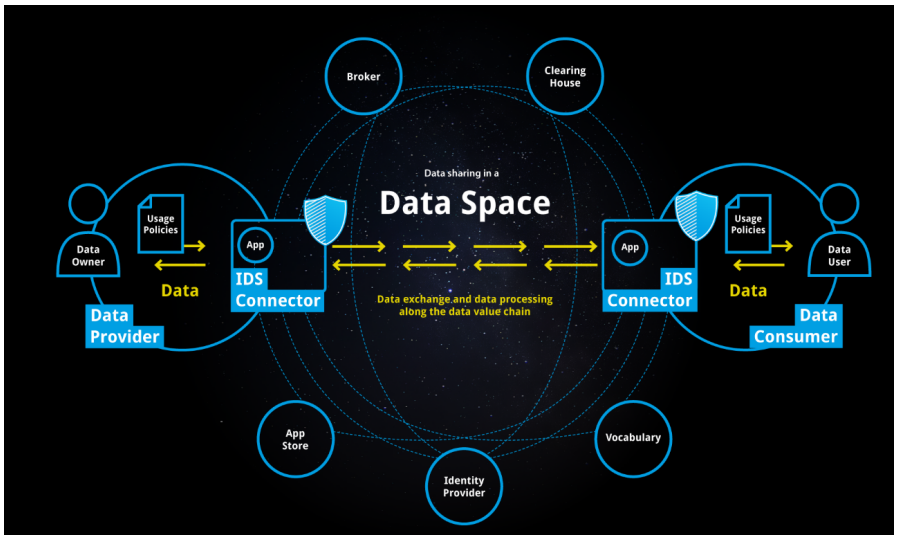

Figure 2.2: Dataspace Components [10]

Connector

The Connector is the core of the IDS architecture and most likely its most critical component. Being the link between an organization's internal systems and the external data space environment, the connector allows participants to keep sovereignty over their data. It enables secure two-way communications, enforces data transaction usage policies, logs system activity and logs all transactions for auditing. Architecturally, the connector separates control and data planes: the former is responsible for managing data, routing and processing, whereas the latter handles actual data exchange [6].

It has several logical layers from a functional standpoint. The Core Container provides core communication services like the Data Router, which manages message flow according to configuration policies, and the Data Bus, which manages local data movement and exchange among the other components. The other containers, App Store container and Custom container, execute certified applications from the IDS App Store or custom tools that the participant installs. Significantly, all processing is done as close to the data source as possible-maintaining locality and minimizing unnecessary exposure [12].

Identity Provider

No secure data space exists without reliable participant authentication. The Identity Provider maintains the trust system by issuing and authenticating identity credentials. These include cryptographic certificate management (X.509v3) and dynamic security token provisioning that inserts ephemeral attributes, such as trust scores or known vulnerabilities. The service is made up of three sub-elements:

- Certification Authority (CA): issuing and revoking certificates
- Information Service (ParIS): for depositing self-descriptions
- Dynamic Attribute Provisioning Service (DAPS): for verifying live participant attributes [6].

In contrast to conventional Public-Key-Infratructure (PKI) based identity infrastructures, IDS decouples identity authentication from the exchange of identity characteristics. Static certificates authenticate the participant, whereas dynamic tokens handle more nuanced, context-dependent traits. This multi-layered system enhances flexibility and allows for the dynamic nature of trust in federated contexts [12].

Metadata Broker

The Metadata Broker is the registry of the data space. It allows Connectors to publish metadata about their available data sources and supports discovery by other actors. In a publish/subscribe mode, the Broker sends notifications to connectors about registered datasets, status changes, or terms of access. It should be noted that the Broker only works with metadata - it does not access, manage, or transfer the underlying data, which still remains the property of the data provider. This architecture enables the idea of data sovereignty with transparent matching between the consumers and providers [12, 6].

Clearing House

The Clearing House provides audit, billing, and settlement facilities for data exchanges. All data exchanges are captured while the transactions are being exchanged, which can later be used to resolve disputes or further analysis. Once the data transaction has taken place, each party logs its details, allowing billing and non-repudiation. Rollback of failed or incomplete transactions is also supported by the Clearing House, adding a critical layer of robustness to federated data sharing [12].

App Store

IDS App Store enables deployment of modular software building blocks - Data Apps - that can extend the Connector with capabilities. The apps may serve a variety of purposes, such as analyzing data, enhancing its semantics, or visualization. The App Store enables publication, querying, rating, and monetizing of apps, enabling open development that enables actors to adapt their Connector environments without compromising basic IDS interoperability [6, 12].

Vocabulary provider

Semantic interoperability is necessary to enable meaningful sharing of data. The Vocabulary Provider manages reference ontologies, domain schemas, and metadata vocabularies used to define data assets. The provider facilitates connectors to understand and label data in standard forms, even across domains. The provider maintains version-controlled vocabularies and supports collaborative development, thus facilitating the technical foundation of discoverability of data, enforcement of policy, and machine-readability within the ecosystem [12].

Together, these key components make up the functional and governance backbone of the IDS model. They enable secure, traceable, and sovereign data exchange,

supporting an extensible, modular federated system aligned with European principles of interoperability and trust.

2.1.5 Roles in a Dataspace

Building on the architectural view, we delineate the roles and organizational responsibilities that enable compliant participation. Participants in federated data spaces adopt roles that fit their duties and responsibilities within the ecosystem of data sharing. To ensure organized interactions, legal clarity and technical interoperability among all parties concerned, these responsibilities are crucial.

Based on their operational duties, extent of involvement in data transactions and connection to shared infrastructure services, participants can be divided into various groups.

Category 1 - Core Participants

These actors are the main participants in any transaction involving operational data and are actively engaged in the data exchange procedures [12].

Data Owner A data owner is either a legal or natural person who holds legitimate authority over a dataset. Ownership entails two primary responsibilities: the ability to define usage conditions (including licensing, usage restrictions and payment models) and the capacity to provide access to the data under those defined terms [12].

Data Provider This role refers to the entity that technically exposes data for consumption by third parties. While often the same as the data owner, the provider focuses on operational delivery -implementing the IDS-compliant infrastructure necessary for secure sharing and optionally enhancing or transforming the data using Data Apps [12].

Data Consumer The consumer is the counterparty to the data provider. Consumers can search for datasets via a Broker Service and then connect to a data provider [12].

Data User A data user holds the right to utilize data under the terms specified by the data owner. While in most scenarios the user and consumer roles overlap, there are cases in which these roles are distinct [12].

App Provider These actors develop and distribute software components (Data Apps) designed to run within the IDS Connector. Applications must conform to the system's architectural standards and may undergo certification to guarantee trustworthiness. They are described using metadata for discoverability and interoperability [12].

Category 2 - Intermediary

This category includes entities that operate the shared services mediating the federation and trust schemes within a data space. These roles are normally assigned to highly trusted parties and include some of the elements mentioned above, namely the Identity Provider, Metadata Broker Service Provider, Clearing House, App Store and Vocabulary Provider. In practice, a single organization may manage multiple intermediary roles or delegate them across independent providers.

Category 3 - Software and Services

This section includes entities that offer infrastructure, platforms or digital services to participants in the data space [12].

Software Provider Creates and distributes software components (not Data Apps) outside the App Store ecosystem. These may be governance, orchestration or analytics tools.

Service Provider Hosts IDS-compliant technical infrastructure on behalf of participants that lack their own deployment capabilities. This can also include value-added services such as data cleaning or semantic enrichment.

Category 4 - Governance Body

These organizations ensure the integrity, evolution and compliance of the data space ecosystem.

International Data Spaces Association (IDSA) A non-profit consortium responsible for defining, maintaining and advancing the IDS architecture and standards. The IDSA operates through dedicated working groups involving stakeholders from industry, research and government [12].

Certification Body and Evaluation Facility These entities certify both the core technical components, such as the connectors and the participants themselves, guaranteeing compliance to IDS standards and fostering trust within the ecosystem [12].

2.1.6 Ongoing Cloud-to-Edge and Dataspace Initiatives

This section proposes an overview on the ongoing European initiatives that operationalise the principles and architectures outlined above across the cloud-to-edge continuum. We begin with NOUS, which provides the reference cloud architecture central to this thesis (see Chapter 3), and then map the landscape of associations, frameworks and sectoral data spaces that complement it. Together, these programmes illustrate how sovereignty, interoperability and trust are being translated into concrete governance models, middleware and production-grade deployments.

NOUS Among the various ongoing programmes in the European landscape, the NOUS project (Next Generation European Cloud Services) deserves immediate attention, since it constitutes the foundation of this thesis and will be analyzed in depth in the chapter 3. It emerges as a Horizon Europe initiative that aims to lay the foundations for a federated European cloud architecture capable of spanning the entire computing continuum, from the edge to the cloud and high-performance computing. Conceived and developed by a multidisciplinary consortium of research institutions, industrial partners and technology providers, NOUS responds to the need for a unified and sovereign digital infrastructure that can serve diverse domains and future technological evolutions [13].

At the core of its vision, NOUS aspires to create a cloud-to-edge continuum that ensures seamless orchestration of resources, scalability, and interoperability across heterogeneous environments. The architecture is designed to provide compute and storage capacities close to data sources while enabling efficient integration with more powerful infrastructures such as High Performance Computing (HPC) centres, and preparing the ground for the inclusion of quantum computing in the longer term [14].

The project's objectives extend beyond technical performance. NOUS places strong emphasis on trust, sovereignty and openness, ensuring that participants retain control over their data and resources while benefiting from a shared infrastructure. It also aims to foster innovation and competitiveness by offering a flexible environment where new services can be developed, deployed and scaled rapidly. By supporting both established providers and emerging actors, NOUS contributes to strengthening the resilience and autonomy of Europe's digital ecosystem [14].

International Data Spaces Association (IDSA) The International Data Spaces Association (IDSA) plays a pivotal role in shaping the conceptual and technical backbone of dataspaces. It is a non-profit consortium uniting industry, academia and government actors to define standards, governance models and architectures for sovereign and secure data exchange [10]. Its flagship contribution is the IDS Reference Architecture Model (IDS-RAM), which formalizes components such as connectors, identity providers, clearing houses and brokers [11]. These elements ensure interoperability, traceability and trust in federated ecosystems. IDSA also oversees certification processes for both technical components and participating organizations, thus guaranteeing adherence to agreed compliance levels. IDSA's importance lies not in providing infrastructure itself, but in creating a framework of trust and sovereignty that other initiatives, projects and platforms can adopt [10]. Without such standardization, the proliferation of dataspaces would risk fragmentation and incompatibility.

FIWARE The FIWARE Foundation is another key actor, focusing on the development and adoption of open standards and open-source software components. FIWARE provides modular building blocks - often referred to as Generic Enablers - that can be used to design smart solutions across domains such as smart cities, industry, mobility or agriculture. Its emphasis on interoperability and transferability allows solutions built in one context to be replicated in another without vendor lock-in. In the context of dataspaces, FIWARE contributes significantly through its APIs, context brokers and data models, which ensure that data and services remain extensible and portable. By defining these standards, FIWARE complements architectural efforts like NOUS, since a cloud continuum without interoperable applications would remain underutilized [15].

Data Spaces Support Centre (DSSC) The Data Spaces Support Centre (DSSC) was created by the European Commission to provide coordination, methodological support and knowledge exchange across the various common European data spaces. The DSSC develops reference models, glossaries, blueprints and guidelines that ensure coherence between sectoral dataspaces such as those for mobility, health, industry, finance and energy. It also maintains a strong link with the SIMPL initiative (Smart Middleware Platform), providing operational middleware services [16].

Gaia-X Gaia-X emerged as a flagship European initiative to counterbalance dependency on non-European cloud providers and to foster a federated and sovereign data infrastructure. Rather than becoming a cloud provider itself, Gaia-X provides a federation framework: a set of rules, labelling mechanisms and services that allow providers to interconnect their infrastructures in a transparent, secure and sovereign manner [17]. Central to Gaia-X are the Federation Services, a toolbox covering identity and trust management, sovereign data exchange, catalogues, compliance frameworks and integration portals. These services enable the creation of sectoral Lighthouse Projects such as Catena-X (automotive), AgrospAI (agriculture) or COOPERANTS (aeronautics), which demonstrate concrete applications of Gaia-X principles [18]. While Gaia-X does not aim to become a cloud platform itself, it sets the conditions of trust and interoperability that enable a European data ecosystem to flourish.

European Open Science Cloud (EOSC) The European Open Science Cloud (EOSC) represents a domain-specific dataspace dedicated to the research community. Its mission is to provide a trusted environment where scientists can find, access, share and reuse data, tools and services across disciplines and countries. Rooted in the FAIR principles (Findability, Accessibility, Interoperability and Reusability), EOSC acts as a collaborative platform that integrates national and thematic

data infrastructures into a single European environment. EOSC illustrates how dataspaces can serve specific communities with tailored governance and service catalogues. By aligning incentives for researchers and ensuring that scientific data remains interoperable and accessible, EOSC fosters innovation while adhering to principles of sovereignty and trust [19].

Sectoral and Industrial Data Spaces

Catena-X Catena-X is the first production-grade dataspace dedicated to the automotive industry. It creates a sovereign and secure environment where manufacturers, suppliers, service providers and regulators can share data along the entire value chain. Its primary objectives include traceability of parts, quality management, circular economy tracking and sustainability monitoring. By applying Gaia-X principles and the IDS-RAM, Catena-X ensures that each participant retains control over its data while enabling interoperability with others [20].

Mobility Data Space The Mobility Data Space (MDS), which is supported by the German government and aligned with Gaia-X, provides a decentralized marketplace for data related to mobility. Both public and private stakeholders publish and use datasets under clearly defined conditions. Applications like multimodal trip planning, traffic optimisation, and public transportation dataset integration are made possible by the MDS, demonstrating how a dataspace can promote innovation in transportation services [21].

Smart Connected Supplier Network The Smart Connected Supplier Network (SCSN) was created in the Dutch high-tech manufacturing sector to strengthen supply-chain resilience, where delays of single components can block entire production lines. Developed by TNO with partners such as KMWE and Brainport Industries, SCSN uses IDS standards to enable seamless data exchange between suppliers, manufacturers and service providers. Participants can share technical product data, track deliveries and maintain visibility without vendor lock-in [22]. After its success in the Netherlands, the initiative is expanding internationally under the "Market 4.0" banner, with pilots in France, Spain and Japan.

Eona-X Eona-X is a dataspace dedicated to mobility, transport and tourism, spanning airlines, rail and bus operators, car and bike services, and hospitality providers. Led by a consortium including Amadeus, SNCF, Air France, KLM and newer partners, it builds a sovereign environment for data sharing to improve travel experiences and operational efficiency. A central catalogue of data and services supports new business models and cross-sector collaboration. Early demonstrators highlight use cases such as intermodal journey planning and enhanced security,

positioning Eona-X as a reference for trusted dataspaces in transport and tourism [23].

The evolution from centralized infrastructures to federated systems, and ultimately to dataspaces, reflects a broader shift in how data is conceived, controlled and valued in the digital economy. Dataspaces apply principles of sovereignty, interoperability and trust based on architectural blueprints such as IDS-RAM, supported by roles and governance arrangements. Their applicability is evident in the variety of European projects and sectoral applications presenting how theory is being translated into working worlds. Dataspaces are not only emerging as a technical paradigm but also as a foundation for Europe's approach towards an economically strong, sovereign and innovation-driven data economy.

Overall, data spaces mark a significant step in data management models. They maintain control over shared resources while facilitating collaboration, data governance, and interoperability across organisational boundaries. This theoretical shift paves the way for investigating how technology infrastructures may foster trust and accountability in distributed settings.

2.2 Blockchain Technology

Blockchain is a digital, distributed ledger that records transactions in units called blocks, which are linked through cryptographic hashes and appended in chronological order. Copies of the ledger are replicated across multiple network nodes, which makes the record tamper-evident and resilient to single points of failure [24]. Blockchain is best understood as an append-only ledger rather than an editable database in the traditional sense, even though the two serve complementary purposes in information systems [25].

Although the technology is often associated with cryptocurrencies, the two are not the same thing. Cryptocurrencies are one application of blockchain, while the underlying ledger can be used to store and transfer many forms of value and state, from digital assets to verifiable data, across different domains [26]. The appeal of blockchain rests on a small set of general properties - decentralization of control, integrity of the recorded history, and transparent auditability - that enable parties to coordinate without relying on a single trusted intermediary [24]

2.2.1 Origins

The roots of blockchain technology are traced back to cryptographic advances that preceded Bitcoin. Whitfield Diffie and Martin Hellman established public-key cryptography in the early 1970s, a breakthrough that facilitated secure communication without shared secrets, laying the groundwork for digital security [27]. David Chaum built on this during the 1980s with the establishment of eCash, an

electronic payment system utilizing blind signatures to maintain confidentiality and security against fraud, revolutionary ideas for secure, anonymous online payments [28]. These developments laid the groundwork for the development of other innovations in decentralized systems. In the early 1990s, Stuart Haber and W. Scott Stornetta introduced digital time-stamping techniques to ensure the electronic documents' integrity. Their approach enabled documents to be neither backdated nor forward-dated, even when the time-stamping service itself was compromized [29]. They employed one-way hash functions and later introduced Merkle trees to aggregate multiple document commitments, enabling efficient verification at scale and introducing the concept of a hash-linked chain of records [29, 30].

These ideas came together with Bitcoin. Satoshi Nakamoto proposed in a white paper in 2008 a peer-to-peer electronic cash system that would prevent double spending without the need for a trusted third party. The system combined a distributed time-stamping functionality, hash-linked blocks, and a proof-of-work model for consensus rewarding good conduct [31]. Its security relies upon the assumption that the honest nodes control most of the network's computational power and thus it is costly for an attacker to manipulate transaction history further back than a few blocks [31]. This approach resolved the Byzantine Generals' Problem, a long-standing issue in distributed systems where the participants must come to a shared truth despite potential dishonesty or failure, with the use of computational effort and transparent verification [31]. Bitcoin's success paved the way for more widespread experimentation, taking blockchain's use beyond basic payments to complex state machines and decentralized applications. It also brought issues of governance, privacy and trade-off between decentralization, security and scalability, otherwise referred to as the scalability trilemma.

2.2.2 Core Principles

Blockchain belongs to the broader class of distributed ledger technologies - DLTs. A DLT is a distributed database organised as a ledger: it maintains an append-only record of state changes, replicated across multiple machines, and kept consistent by a protocol rather than by a single operator. Blockchain is one specific family of DLTs that represents the ledger as a hash-linked sequence of blocks, each block committing to its predecessor so that altering past records becomes practically infeasible without redoing subsequent work and subverting the protocol's governance assumptions [26].

Distributed and *decentralized* are related but distinct concepts. A DLT, specifically the blockchain, is *distributed* in the sense that copies are replicated across sites. Degree of decentralization depends on governance and admission rules, which in turn shape properties such as openness, auditability and fault tolerance. A blockchain is *architecturally decentralized*, meaning no single point of failure exists,

but is *logically centralized*, since only a single logic status is admitted and it depends on which protocol is applied by the nodes to agree on the same logic status, as defined in section 2.2.5.

On the authority level, three models are identified:

Permissionless blockchains are open networks that allow anyone to read the ledger, submit transactions, and attempt to publish blocks without prior authorization. These systems use consensus mechanisms that require participants to expend or escrow scarce resources when proposing blocks, such as Proof of Work or Proof of Stake outlined in Section 2.2.5, to discourage attempts to undermine ordering or validity because openness suggests the presence of potentially adversarial actors. Incentives are integral: publishers of protocol-conforming blocks are typically rewarded with a native cryptocurrency, aligning individual behavior with network rules. Because participation rights are not gated, the ledger is broadly readable and writable, and software implementations are often released as open source [26].

Permissioned blockchains restrict who may publish blocks through an admission process governed by a designated authority or consortium. Read access and the ability to submit transactions can also be limited to authorized parties, and implementations may be open or closed source. Despite access controls, these systems retain the salient properties of ledgers - traceability of asset transfers and resilient, replicated storage - while using consensus protocols that are generally faster and less resource-intensive because participants are identified and subject to revocation. This model suits organizations that require tighter control or that collaborate without full mutual trust. It enables stronger operational transparency, including continuous audit by oversight entities, and supports selective disclosure in which transaction details are visible only to credentialed parties. Where identities are mandatory, actors are not anonymous or even pseudonymous, which creates natural disincentives for fraud and facilitates recourse through established legal remedies if misconduct occurs; however, if a single entity controls admission, users must ultimately trust that authority [26].

Consortium blockchain is a permissioned network governed by a group of independent organizations that jointly control validator selection, membership admission, and protocol configuration. It is a hybrid solution between the public model and the completely private one, where write permissions are restricted to preapproved validator nodes operated by the members, while read permissions may be public or restricted [32]. Consensus typically uses identity-aware, low latency protocols, relying on revocation and legal accountability rather than resource expenditure.

2.2.3 Structural Elements

Addresses

An address is a special identification - specifically an alphanumeric sequence of characters - used in blockchain transactions that identifies the sender or recipient of a transaction. A public key or a value obtained from a public key are the two most common types of addresses. Addresses are more concise than the public keys and are not confidential. A technique for creating an address is to begin from a public key, apply a cryptographic hash function to it and finally transform the hash into text. In order to improve privacy and security, users frequently create a new, distinct address for every transaction, even though addresses are technically reusable by the same user and are intrinsically unique [24].

In numerous blockchain networks, especially permissionless ones, users are directly tasked with handling and protecting their private keys. These keys serve as cryptographic credentials that permit transactions and manage access to digital assets. To minimize the risks associated with manual storage, users generally depend on specific applications referred to as *wallets*. A wallet can hold both private and public keys, handle linked addresses and typically offers extra functionalities like balance tracking and transaction records. The protection of private keys is crucial. If a key is missing, all assets associated with it become permanently unreachable, as it is computationally impractical to regenerate the identical key. On the other hand, if a key is compromised, the intruder obtains complete access to the assets it secures. Consequently, certain users choose hardware-based secure storage options, while others prefer specialized custody or escrow services [26].

Transactions

A transaction represents the atomic unit of state change, a record of an event. It encodes inputs, outputs and validity data, and is authenticated by digital signatures. Valid transactions propagate over the network, await ordering, and once confirmed become part of the durable record. In permissioned systems, submission and validation rights are scoped to authorised participants; in permissionless systems, any node may relay candidate transactions subject to protocol rules [26].

Blocks

Each block is an aggregation of transactions that are logically bundled together. The structure of each block depends on the type of blockchain. Generally, it is divided into the *block header* and the *block body*. The former contains the following components:

- *Pointer to the previous block's hash* in the chain: it is a fixed-length digest that

creates the one-way link that chains blocks in chronological order. Any change to a past header or its contents alters its hash, which in turn invalidates the pointer stored by all descendants. This property makes the history tamper evident and raises the cost of rewriting it as the chain grows [24].

- *Timestamp*: records when the block was created. Rather than an external source of truth to the system, it represent a consistency signal for the ordering layer.

- *Nonce*: it is a number only used once, often used in cryptographic operations to ensure replay protection, authentication and secure encryption. It can be integrated with data to generate various hash digests for each nonce - hash (data + nonce) = digest [26]. Merely altering the nonce value offers a way to achieve various digest values while retaining the same information. This method is employed in the proof of work consensus mechanism.

- *Merkle root*: a Merkle tree is a structure that compactly commit to large transaction sets. The digest associated to a Merkle tree - Merkle root - is included in the block. Any modification to a transaction or to its position in the tree changes the root. This structure enables efficient inclusion proofs: a verifier can check that a transaction is part of the block by inspecting only a logarithmic number of sibling hashes rather than downloading the full block, which is essential for light clients [31].

The *block body* instead carries the ordered set of valid transactions that the network has accepted for inclusion under the protocol's rules. Each transaction encodes inputs, outputs and authorization data and, once confirmed, effects a state transition on the ledger [24].

As represented in Figure 2.3, the structure explained above is presented.

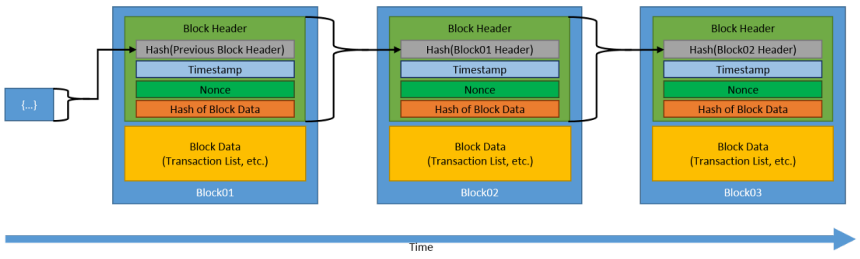

Figure 2.3: Generic Chain of Blocks [26]

Peer-to-peer network

Blockchains operate on top of P2P networks. Peers find one another, spread blocks and transactions, and maintain partial views sufficient for eventual consistency [26]. The absence of a central message broker removes single points of failure but places greater emphasis on protocol-level validation and anti-abuse measures.

Nodes and roles

A node is a single system that takes part in a blockchain network, and contributes to its functionality.

The *full nodes* store and verify the entire chain, enforce protocol rules, and transmit data. Since it stores the entire ledger, it plays a critical role in ensuring the network's integrity and security. The *publishing node* instead, is a type of full node that, in addition to store the blockchain and validate transactions, creates and distributes new blocks to the network. A *lightweight node* differs from a full node since it does not store a copy of the blockchain. As a consequence, it makes it less resource-intensive because it relies on full nodes to process and validate its transactions. However, it remains dependent on other nodes for operation.

Other roles depend on the specific system, miners in proof-of-work or validators in proof-of-stake (detailed in Section 2.2.5) [26].

2.2.4 Cryptography

Cryptography is the foundation of blockchain security, enabling integrity, authenticity, and confidentiality in a decentralized environment where there is no central authority imposing trust. There are two main families of cryptographic primitives that underlie blockchain systems: *hash functions*, which ensure data integrity and immutability, and *asymmetric-key cryptography*, which enables authentication and authorization of transactions.

Cryptographic hash functions

A cryptographic hash function is a mathematical function that takes input data of any size and returns a fixed-size output, commonly referred to as a *digest*. These functions possess three significant properties: preimage resistance (it is computationally infeasible to reconstruct the original input from its hash), second preimage resistance (it is infeasible to discover a different input with the same digest), and collision resistance (it is infeasible to find any two distinct inputs with the same output) [26].

Hash functions serve several purposes in blockchains: they link blocks to one

another through hash pointers in the block header, they commit to all the transactions within a block efficiently through Merkle trees, and they are used to derive unique identifiers such as wallet addresses. Any change to prior data changes its hash, thus making following chain links invalid. It is this structure that provides tamper-evidence, which provides the practical immutability characterizing blockchain ledgers. Standards that are widely used are SHA-256 (Secure Hash Algorithm), and Keccak. The SHA-256 has an output size of 256 bits, represented by a 64-character hexadecimal string (e.g., the input text "Hello, World!" corresponds to the following SHA-256 digest:
0xdffd6021bb2bd5b0af676290809ec3a53191dd81c7f70a4b28688a362182986f).

Asymmetric-key cryptography

Blockchains also rely on asymmetric-key cryptography, also referred to as public key cryptography. It is a system that utilizes a mathematically related pair of keys: a public key, which can be safely divulged, and a private key, which must stay confidential. Although mathematically related, it is computationally infeasible to derive the private key from the public key [26].

The mechanism enables mutually untrusted participants to interact securely digitally. A user's private key digitally signs transactions, forming a cryptographic proof that only the legitimate owner of the said key could have possibly initiated the operation. The signature is verified by other nodes using the corresponding public key. The ability to verify signatures without ever revealing the private key itself is what enables both the authenticity and non-repudiation of transactions [24].

On a practical level, public keys also play a foundational role for blockchain address derivation, where private keys act as authorization credentials to unlock associated assets. When sending a transaction, a user's private key signs transaction data, and the network nodes verify that signature against the derived public key, before committing the transaction to the ledger. Through this mechanisms, only valid private keys can execute valid transfers, and everyone can independently check the validity of the operations [24].

While asymmetric-key cryptography forms the foundation of establishing trust in permissionless environments, it is computationally more expensive than symmetric-key cryptography, which uses the same shared secret for decryption and encryption. Symmetric schemes, such as Advanced Encryption Standard (AES), are orders of magnitude faster but require communicating parties to share a key in advance - a limitation in open, decentralized networks where there is no pre-existing trust [24].

To conclude, asymmetric cryptography plays four important roles in most blockchain designs:

- Transaction signing – Private keys are used to digitally sign transactions,

authenticating the sender.

- Address derivation – Public keys are mathematically translated into blockchain addresses, which serve as identifiers for users.
- Signature verification – Public keys are used by nodes to verify that a transaction signature corresponds to an authentic private key owner.
- Proof of ownership – The ability to sign a valid transaction is the cryptographic proof that the signer possesses control of the assets in question.

In permissionless blockchains, users will more likely generate and manage their own key pairs themselves. Permissioned blockchains can, however, make use of existing Public Key Infrastructures (PKI), such as corporate directory services, to manage credentials more centrally [26].

2.2.5 Consensus Algorithms

Consensus algorithms are the backbone of blockchain systems because they enable an untrusted network of nodes to come to an agreement on a consistent ledger state. In contrast to centralized databases, where order is ensured by a trusted third party, blockchains implement consensus in order to provide safety, liveness, and fault tolerance in adversarial settings. The goals of a consensus protocol are: (i) to get all honest nodes to agree on the same sequence of blocks, (ii) to commit just valid transactions, (iii) to tolerate malicious/faulty nodes, and (iv) to guarantee progress under fair conditions [24].

Below the technical activity of the two most common paradigms, Proof-of-Work and Proof-of-Stake, found in permissionless systems is analyzed.

Proof-of-Work (PoW)

Proof-of-Work is the earliest consensus mechanism to be applied in Bitcoin [31]. It relies on having nodes (miners) compute computationally demanding puzzles: i.e., they iterate nonces and calculate a hash of the block header such that the calculated hash is compliant with a global difficulty constraint (e.g., a number of leading zeros). The first successful miner to identify a valid nonce broadcasts the block; it is inexpensive for other nodes to verify that the hash is valid without repeating the mining procedure [24].

Because block creation is probabilistic, the chain selection rule is the "longest chain" or "most work" rule: nodes mine the branch with highest total proof-of-work [26]. Any other attacker attempting to reverse history must possess more than 50 % of the network hash power in order to build an alternate chain faster than honest miners - a so-called 51 % attack. From a technical perspective, the

longest-chain rule PoW-based consensus provides some of the best consistency and adversarial reorganization resistance guarantees, provided the adversaries have subthreshold power. However, PoW suffers from fundamental inefficiencies: high energy consumption, low throughput (due to block size, propagation delay and confirmation), and centralization pressure (due to scale economies in mining) [24].

Furthermore, PoW algorithms require dynamic difficulty adjustment: the levels of difficulty must adapt continuously in order to maintain stable block intervals despite changes in total hash power. This temporal coupling can lead to instability when hash power fluctuates rapidly [24].

Proof-of-Stake (PoS)

Proof-of-Stake replaces computational work with economic stake as the scarce resource. Validators lock up some tokens to gain the right to propose or validate blocks and to be eligible for rewards proportional to their stake. As no vast amount of computation is involved, PoS reduces energy consumption substantially and speeds up finality in most designs [24].

Existing PoS systems incorporate slashing penalties: when a validator behaves maliciously (e.g., signs incompatible blocks), some or all of their stake is fined. This creates an inherent cost to misbehaviour. Other than raw chain weight, PoS protocols use alternative fork-choice rules.

There are several PoS architectures:

- Chain-based PoS: validators propose and extend blocks similar to PoW, but selection is stake-proportional.
- Committee-based or BFT-style PoS: validators are formed into committees and conduct voting rounds to finalize blocks (e.g., Tendermint, HotStuff).
- Delegated or hybrid PoS: in Delegated PoS (DPoS), the larger community delegates stake to a smaller set of validator nodes; or hybrid PoW/PoS which uses both mechanisms for block proposal and settlement [24].

But PoS is not without downsides:

- the nothing-at-stake problem: being computationally inexpensive, validators can sign multiple forked chains, unless slashing penalizes equivocation.
- Long-range attacks: validators can attempt to build alternative chains from deep back, especially if old keys or stakes are made worthless; most chains mitigate this through checkpointing or finality gadgets.
- Economic threats: freshly emerging use of DeFi, staking derivatives, or lending reduces effective stake security because validators will move value out of staking to seek better returns off-chain.

- Network-layer attacks: certain recent research suggests that PoS networks face routing manipulations, delays, or censorship at the communications level, which can influence block proposals and slashing.

- Decentralization trade-offs: exacting high minimum stake or validator requirements can enhance performance or security, but reduce the size of participants, tilting in favor of centralization [26].

From the formal security point of view, PoW's longest-chain rule ensures consistency and probabilistic finality under honest-majority assumptions, while PoS can reach commensurate amounts of security only if safety and liveness trade-offs are balanced.

2.2.6 Smart Contracts

The concept of smart contracts was first articulated by Nick Szabo in the mid-1990s, notably in 1994, as computer-based transaction protocols designed to enforce contractual clauses while reducing reliance on trusted intermediaries [33]. Initially, the notion was closely tied to the automation of legal contracts, with the aim of reducing transaction costs and dependence on intermediary entities by guaranteeing that obligations were automatically fulfilled once specific conditions were satisfied. Szabo's idea established the foundation for subsequent applications of self-executing agreements on distributed ledgers. Indeed, with the advent of blockchain, the definition has shifted: today, smart contracts are commonly understood as tamper-proof computer programs that autonomously update the state of a distributed ledger [34]. In this sense, the term "contract" is somewhat misleading, as these programs are not necessarily legal agreements but deterministic code running across decentralized nodes.

From a technical standpoint, a blockchain can be described as a finite state machine, and smart contracts represent state-transition functions that map a given state and input transaction into a new state. This characterization highlights two essential aspects: smart contracts must modify the system's state to produce verifiable results, and they must do so deterministically [34]. Determinism ensures that identical inputs always yield identical outputs across all nodes, enabling consensus [26]. Without determinism, inconsistencies across replicas would undermine the ledger's integrity. This property also explains why functions that merely read data or perform computations without altering the blockchain state cannot be considered smart contracts in the strict sense [34].

The execution of smart contracts exhibits several distinctive properties. First, they are tamper-proof in the sense that altering their execution requires collusion by a majority of nodes, which is generally infeasible in well-decentralized networks [34]. Second, their outputs are verifiable at any point in the future, a requirement

that prevents inconsistencies when verifying execution retrospectively [34]. Third, while often described as immutable, smart contracts are only immutable in a relative sense: blockchains are append-only ledgers, which means that new versions of contracts can be deployed, enabling contract upgradability [34]. Mechanisms such as proxy contracts in Ethereum demonstrate how a contract can delegate execution to updated code while retaining the same address, illustrating the nuanced relationship between immutability and flexibility [26, 34].

Another important aspect is that multiple implementations can belong to the same equivalence class provided they produce the same state transitions. Nevertheless, the usual practice of creating a smart contract once and deploying it to all nodes is counter to the nature of blockchain, which presumes that the nodes do not trust one another. Thus, a more rigorous policy would have every node develop or, at least, validate its own implementation, so that correctness should never be assumed. Although redundancy in such a manner enhances immunity to possible bugs, it has managerial implications which are high, as development costs are duplicated among participants instead of pooled, hence raising the general economic cost of blockchain-based projects [34].

Despite their potential, smart contracts face significant technical challenges. One critical issue is the management of external data. Since smart contracts cannot directly access off-chain information, they rely on oracles - trusted third parties that feed external data into the blockchain. They act as reliable data providers, yet this dependence compromises a focal promise of blockchain, decentralization. Bringing trusted third parties back into the system, diminishes resilience and introduces new vulnerabilities [34]. This reliance creates the *garbage in, garbage out* problem: while blockchain ensures data integrity once stored, it cannot guarantee the correctness of incoming data [35]. Consequently, decision-makers need to thoroughly assess the structure of data collection processes and the reliability of oracle networks to reduce GIGO risks [35].

Smart contracts are also exposed to vulnerabilities stemming from coding errors and design flaws. High-profile incidents such as the Decentralized Autonomous Organization (DAO) hack have highlighted issues like reentrancy attacks, transaction-ordering dependencies, and exception handling problems. For this reason, formal verification, static analysis, and independent auditing have become critical components of smart contract development [36].

2.2.7 The Scalability Trilemma

The execution of smart contracts is fundamentally constrained by the scalability issues of blockchain infrastructures. In public blockchains like Ethereum, every transaction needs validation from all active nodes, ensuring decentralization and resistance to tampering but limiting overall throughput. Even in permissioned

systems intended for industrial uses, it's essential to balance high transaction throughput with the decentralization of trust [37]. Although permissioned systems can achieve better scalability by limiting participation and refining consensus algorithms, they might sacrifice the level of openness and resilience inherent in fully decentralized networks.

This trade-off is frequently represented through the scalability trilemma, which indicates that a blockchain system finds it challenging to optimize three characteristics at once: decentralization, security and scalability. In practice, systems need to emphasize two dimensions while sacrificing the third. Public blockchains attain significant decentralization and security yet encounter scalability issues, while private or consortium blockchains enhance scalability by restricting participation but could compromise decentralization [38]. This trilemma directly impacts smart contracts, as their deterministic execution among all validating nodes is resource-heavy and consequently exacerbates scalability limitations [34].

In industrial supply chains environments, these problems imply specific operational difficulties. Consequently, only specific data and significant state changes can be recorded on-chain, while ordinary data are stored off-chain to maintain scalability. This targeted method emphasizes the principle according to which blockchains and smart contracts should ensure the integrity and traceability of the most critical data, instead of functioning as stores for all information.

In summary, blockchain establishes a decentralized environment in which trust is the result of cryptographic guarantees and public verification rather than central authority. Allowing for transparency, integrity, and verifiable record-keeping, it creates a strong foundation for reliable data operations in advanced digital ecosystems. These properties become more relevant in contexts where large amounts of data are generated and acted upon in real time.

2.3 Edge Computing and Smart Cities

2.3.1 Smart Cities and the Evolution of Intelligent Mobility

The concepts introduced in the previous sections - federated data management through dataspaces and trust mechanisms enabled by blockchain - find concrete application within the context of smart cities and smart mobility. These domains represent complex, data-intensive ecosystems where multiple actors continuously produce, exchange, and consume information across organizational and technological boundaries. As such, they embody an ideal environment to observe how decentralized infrastructures and edge computing can enhance interoperability and accountability in data-driven urban services.

In the last few years, urban ecosystems have undergone a revolutionary transformation driven by digitalization, data-driven analytics, and ubiquitous connectivity. The smart city has emerged as a unifying vision for the integration of information and communication technologies (ICT), Internet of Things (IoT) devices, and advanced analytics to optimize city operations, enhance sustainability, and improve citizens' quality of life. A smart city is characterized by its capacity to collect, process, and analyze enormous volumes of heterogeneous data generated by its infrastructure - energy networks, public services, and transportation networks - to support data-driven decision-making and automation [39]. The transportation sector particularly has been at the forefront of this revolution, with the advent of the concept of smart mobility, involving intelligent, interconnected, and green transportation systems.

Smart mobility represents the confluence of connected vehicles, advanced sensing capabilities, wireless communication, and artificial intelligence (AI) with the aim to create adaptive and responsive transport systems. They are designed to reduce congestion, emissions, and enhance the safety and efficiency of mobility in densely populated cities. Intelligent Transportation Systems (ITS) form the operational backbone of smart mobility initiatives, integrating real-time data from vehicles, road infrastructure, and environmental sensors to manage and optimize traffic flow [39]. The evolution from conventional ITS to connected, automated mobility has been enabled by the maturation of enabling technologies such as Vehicle-to-Everything (V2X) communication, 5G networks, and more recently edge and cloud computing infrastructures [40].

In the vision of the smart city, transport is no longer viewed as an independent subsystem but as part of an integrated urban ecosystem. Sensor data from roads, intersections, and vehicles is aggregated and analyzed to coordinate traffic lights, provide predictive analytics for congestion management, and direct emergency response operations. Real-time management of urban mobility represents the transition from rigid, centralized control to adaptive, distributed systems enabled by edge computing and next-generation networks [41].

2.3.2 Safety and the Protection of Vulnerable Road Users

A central component of smart mobility is enhancing road safety, particularly for Vulnerable Road Users (VRUs) - pedestrians, cyclists, and motorcyclists - who suffer a disproportionate percentage of traffic accidents. According to the European Transport Safety Council, VRUs account for nearly half of all road fatalities globally. The multimodality and density of city mobility accelerate risks, and thus the integration of digital safety features becomes an essential imperative of smart cities [39].

Traditional vehicle-based safety systems, such as anti-lock braking and adaptive

cruise control, possess limited situational awareness that hardly extends beyond the immediate surroundings of the vehicle. Intelligent V2X communication systems, nonetheless, widen perception by allowing vehicles to communicate with surrounding infrastructure (V2I), other vehicles (V2V), pedestrians (V2P), and wider networks (V2N). Such extensive connectivity supports cooperative awareness messages that can alert vehicles to the presence of VRUs, road hazards, or potential collisions prior to their occurrence [40].

For instance, V2P applications allow the transmission of a pedestrian's location and motion parameters from a smartphone or wearable device they are carrying to surrounding vehicles, which can then utilize predictive collision avoidance algorithms. Intersections fitted with sensors and cameras can also sense the presence of a pedestrian or cyclist and broadcast adaptive signal priorities to approaching vehicles. These sensor data are processed locally by edge-enabled computing infrastructures to generate real-time safety alerts and hence decrease latency - a critical factor where milliseconds can be the difference between an accident being prevented [39].

By integrating edge analytics and AI-based perception, smart mobility networks extend their protective role to all road users, and not just occupants of vehicles. This sort of convergence of connectivity, sensing, and computing then represents a paradigm shift from a reactive to proactive safety management strategy, aligning with the overall ambitions of Vision Zero initiatives for the elimination of road fatalities in urban mobility.

2.3.3 Multi-Access Edge Computing: Concept and Capabilities

As cities adopt increasingly data-hungry mobility use cases - from connected traffic lights to autonomous driving - the limits of centralized cloud infrastructures manifest. Forwarding massive amounts of sensor data to distant data centers incurs latency, bandwidth consumption, and potential reliability issues. Multi-access Edge Computing (MEC), or Mobile Edge Computing, has therefore emerged as an enabler of next-generation smart transportation infrastructures.

MEC brings computing capacity closer to the data source by inserting processing nodes at the edge of the network - i.e., at base stations, roadside units (RSUs), or even on vehicles themselves. By offloading computation from the cloud to edge nodes, MEC drastically reduces end-to-end latency, enabling real-time analytics and decision-making [40]. In vehicular applications, this architectural decentralization supports latency-critical functions such as cooperative collision avoidance, lane changing, and dynamic traffic signal control.

Edge computing serves as an intermediate layer between cars, sensors, and centralized cloud services. IoT device- and on-board unit (OBU)-sensed data are pre-processed at the edge to eliminate redundant information, detect objects, or

create immediate control actions. Aggregate or contextualized data are uploaded to the cloud for long-term storage or large-scale analytics [39]. This hierarchical structure offers effectiveness and scalability in managing the exponentially increasing volumes of data that are generated by modern transportation infrastructures.

High Mobility Edge Computing (HMEC) is a practical widespread example of MEC, where computation and caching are dynamically coordinated between mobile nodes such as vehicles and drones. HMEC enables distributed decision-making across vehicular networks with robust situational awareness even under high-mobility. The integration of 5G networks into MEC systems extends such capabilities, as 5G provides ultra-low latency and massive device connectivity, both critical enablers of cooperative autonomous driving [40].

Besides, MEC facilitates upper-level AI tasks at the edge, including computer vision, trajectory prediction, and anomaly detection. For instance, an MEC node roadside can process video streams from AI-enabled cameras in real time to detect traffic violations or abnormal pedestrian movement. Edge processing minimizes data exposure risks and ensures the operation of critical safety functions despite a loss of connectivity with the central cloud [41].

2.3.4 Edge Computing in Smart City Infrastructures

The integration of MEC in smart city infrastructures is a tactical step forward in city computing. Traditional smart city platforms relied on the centralization of data centers for the collection and processing of data from the different subsystems - utilities, transport, surveillance, and public services. However, increasing device density along with the stringent latency requirements of mobility applications necessitates the shift towards distributed, edge-centric architectures [39].

In transportation networks, edge nodes are positioned close to sensors embedded in roads, intersections, and cars. Nodes perform functions such as image processing, traffic density analysis, and environmental monitoring. In Intelligent Traffic Light (ITL) systems, for example, cameras and radar sensors feed real-time input to local edge processors that dynamically adjust signal timing to optimize flow and prioritize emergency vehicles. These systems were discovered to reduce travel time by up to 25% and 15–20% of emissions within urban areas [39].

This architecture also supports Virtual Traffic Lights (VTL) - decentralized signaling systems that eliminate physical infrastructure for residential or low-traffic zones. MEC nodes synchronize virtual light projections or in-vehicle notifications, allowing vehicles to autonomously negotiate intersection passage based on real-time situational awareness. This solution improves traffic fluidity, reduces idle time, and supports the safety of drivers and pedestrians [39].

From a systems perspective, edge computing provides the computing substrate for ITS to concatenate VANETs, IoT sensors, and AI-driven control mechanisms.

VANETs leverage V2X communication protocols to exchange information among vehicles and between vehicles and infrastructure. Edge servers in RSUs gather and analyze the information to detect patterns, predict traffic, and choreograph cooperative maneuvers. The result is a multi-layered adaptive network for enabling real-time decision-making between different entities - vehicles, traffic control centers, and pedestrians [40].

Besides this, edge nodes enable mobility prediction, namely the capability to predict vehicle trajectories and traffic flow using historical and real-time data. Proactive signal control optimization, lane assignment, and resource allocation are enabled by predictive analytics executed at the edge. By foreseeing congestion or potential collisions, mobility prediction enhances not only efficiency but also supports sustainability through reduced fuel consumption and emissions [39].

2.3.5 The Synergy of Edge Computing, 5G, and V2X Networks

The convergence of MEC, 5G connectivity, and V2X communication forms the technology foundation of future smart mobility ecosystems. 5G networks provide the ultra-reliable, low-latency communication (URLLC) that is demanded by time-sensitive vehicular use cases, and MEC ensures data processing occurs close to the point of origin. Together, they enable a new class of distributed, cooperative mobility services beyond the constraints of legacy ITS infrastructures [40].

In V2X systems, data interchange and security policies are ensured by edge nodes, which act as brokers between vehicles and the communications network. As several vehicles converge on an intersection, edge nodes can aggregate their telemetry, derive collision probabilities from AI algorithms, and broadcast synchronized alerts in microseconds. It is clear that this use case requires deterministic latency and high bandwidth, both possible through 5G-enabled MEC architectures.

Besides, 5G's network slicing allows for the creation of tailored virtual networks for specific mobility services. V2X safety applications can operate on ultra-reliable slices with guaranteed bandwidth and latency, while less critical setups are utilized by infotainment or navigation services. Edge computing orchestrates such slices dynamically, with optimized resource allocation in heterogeneous urban traffic environments [41].

The addition of blockchain technology - also explored in recent ITS research - extends this architecture by introducing decentralized trust and data integrity to vehicle-to-vehicle communication. Blockchain-enabled MEC nodes can notarize event data such as collision warnings or sensor readings, rendering them immutable and verifiable in distributed transportation networks [41]. This integration of edge computing, 5G, and blockchain forms a solid foundation for transparent and secure smart mobility ecosystems.

2.3.6 Applications and Real-world Implementations

Several cities and industrial initiatives have demonstrated the physical impact of MEC and associated mobility solutions. Pilot deployments in Singapore, Copenhagen, and Dubai have included edge-based traffic management systems that utilize data from roadside sensors and cameras to optimize flow and reduce congestion [39]. In these cases, edge nodes are utilized as local control points, hosting AI algorithms that adjust signal timings or reroute vehicles in response to current conditions.

In Europe, the Cooperative Intelligent Transport Systems (C-ITS) framework is an excellent instance of large-scale coordination among governments, automakers, and telecommunication providers for the standardization of edge-supported vehicular communication. Projects such as the European C-Roads initiative and the U.S. Connected Vehicle Pilot Program have already demonstrated how distributed edge nodes and 5G connectivity can reduce collision rates, speed up emergency response times, and attain improved energy efficiency [40]. Automobile manufacturers have also embraced edge-centric architectures by integrating on-board edge processors that communicate with external MEC nodes to execute cooperative awareness and predictive navigation. These automobiles practice vehicular platooning, a use case where multiple automobiles travel in synchronized convoy using V2V communication and edge-enabled coordination to minimize aerodynamic drag and fuel consumption [40].

At the infrastructure level, low-power IoT sensors are employed to detect vehicle occupancy in edge-enabled smart parking systems. Availability information is communicated to drivers in real time, reducing time spent searching and, therefore, emissions. Similarly, intelligent street lighting systems equipped with motion and environmental sensors use edge processing to dynamically control brightness based on traffic and pedestrian activity. This advantages both energy efficiency and safety [39].

2.3.7 Challenges and Future Directions

Despite its revolutionary potential, the deployment of MEC in smart mobility ecosystems is faced with several technical and organizational challenges. One of the most critical ones is scalability, since cities have to manage thousands of distributed edge nodes with consistent performance and interoperability. Vendor diversity in hardware, communication protocols, and software frameworks complicates integration and standardization [41].

No less critical are security and privacy. The edge computing decentralization is accompanied by a larger attack surface as a large number of nodes handle sensitive data such as vehicle trajectories and personal identities. To thwart these issues, recent research studies have proposed blockchain-enabled cross-domain interaction systems, which provide secure identity authentication and encrypted data sharing

among heterogeneous networks [41]. Identity-Based Encryption (IBE) systems, coupled with decentralized consensus protocols, can enable trustless communication between vehicles, infrastructure, and administrative entities.

Energy efficiency is also an ongoing research direction. On the one hand MEC enables lower data transmission overhead; on the other hand, employing a large number of distributed servers increases local power consumption. Investigations into various methods such as energy-aware orchestration algorithms and renewable-energy-powered RSUs are underway to make edge computing compatible with the sustainability goals of smart cities [39].

On the policy-making level, to implement MEC-based mobility solutions, cross-industry collaboration among municipalities, telecommunication operators, and private enterprises is required. The standardization for V2X communication, edge orchestration, and data governance is essential for ensuring interoperability [40].

Thus, edge computing represents an essential building block for the smart mobility ecosystem: through its application, cities can process, analyze, and efficiently use massive data streams in real time. MEC bridges the gap between connected vehicles, IoT devices, and centralized cloud infrastructures by decentralizing computation. The combination of MEC, 5G networks, V2X communication, and blockchain empowers the realization of a secure, efficient and scalable ecosystem. Moving forward, edge computing and smart mobility will enable the future of intelligent transportation systems, as not only connected and autonomous infrastructures but also sustainable and resilient.

The intersection of the three areas of technology presented in this chapter - dataspaces, blockchain and edge computing - represents a pivotal step towards realizing decentralized and sovereign data ecosystems. Dataspaces provide the governance and interoperability model that enables sovereign data sharing among organizations. Blockchain provides the trust infrastructure required to provide immutability and accountability of such data exchange, while edge computing provides computational proximity required to process data and generate data in real time in dynamic ecosystems such as smart cities.

Their convergence makes possible the secure and transparent flow of data from local perception to cross-organizational valorization, combining technical, semantic, and organizational interoperability. This synthesis underpins the main point of the thesis: that blockchain-based traceability and notarization mechanisms can be applied to enhance auditability and trust in the lifecycle management of edge-generated data in federated dataspaces.

Chapter 3

The NOUS Project

3.1 Cloud Computing and the European Data Strategy

The strategic motivation behind the NOUS project cannot be completely appreciated without first comprehending the European Union's larger goals regarding cloud computing, data infrastructures, and digital sovereignty. Through its Digital Decade and European Data Strategy initiatives, the European Commission has developed a multifaceted framework to assist in the development of cloud infrastructures that are safe, compatible, and sustainable. These regulations are intended to guarantee that European companies, researchers, and government agencies have access to top-notch cloud and edge services that are both technologically sophisticated and in line with EU norms and values. The changing dynamic between distributed edge computing systems and centralized cloud infrastructures is a key component of this approach. While data processing historically occurred in large-scale, centralized data centers, current technological trends indicate a significant transformation: it is projected that, by 2025, 80% of all global data will be processed locally on smart devices at the network's edge [42]. Thus, greater responsiveness, lower latency and more efficient use of energy resources are needed to reflect the shift toward edge computing. It also underlines the importance of ensuring data is stored and transmitted securely across both cloud and edge environments. The concept of a seamless computing continuum, in which cloud and edge infrastructures operate in concert, is foundational to achieving these goals. The EU recognizes that only through the coordinated deployment of energy-efficient and trustworthy computing nodes can the full potential of data-driven innovation be realized.

These infrastructural goals are embodied in two quantitative targets set forth in the Digital Decade policy: by 2030, 75% of European enterprises should be employing cloud-edge technologies in their operations and at least 10,000 highly

secure, climate-neutral edge nodes should be operational across Europe [42]. These nodes are expected to support real-time processing and rapid data transfers, thereby facilitating a host of new services and applications. While the adoption of cloud services among businesses is rising - reaching 45.2% in 2023 according to Eurostat - the uptake remains uneven across business sizes [42]. Large enterprises exhibit higher adoption rates, while small and medium enterprises continue to lag, underscoring the need for policy interventions and support mechanisms.

To support this transition, the European Commission is deploying a comprehensive suite of initiatives that span legislation, funding and infrastructure development. Among these is the anticipated Cloud and AI Development Act. This legislative proposal seeks to triple the EU's data center capacity over the next five to seven years, thereby addressing both the scalability and the energy demands of modern cloud services. The Act will simplify the processes for data center deployment by identifying suitable sites and accelerating permitting procedures for projects that align with the EU's sustainability and innovation criteria. Additionally, the regulation will promote the integration of advanced energy management technologies to improve operational efficiency [43].

In tandem, the Act will establish a single cloud policy across EU Member States for public administration. This will prioritize the adoption of European cloud services for critical use cases, thereby reinforcing digital sovereignty and fostering a competitive cloud ecosystem [43]. The EU's approach is deliberately multifaceted, recognizing that realizing the full benefits of the data economy requires coordinated action across several policy domains - including the Digital Strategy, Industrial Strategy, and major funding instruments such as Horizon Europe, the Digital Europe Programme, and the Connecting Europe Facility.

At the core of this vision is the creation of a European single market for data. This digital single market is intended to enable data to flow seamlessly across sectors and borders while respecting European norms on privacy, data protection and competition. The European Data Strategy seeks to empower consumers, businesses, and public entities to access, share and reuse data under transparent and equitable conditions. The Commission envisions a future where users maintain control over their data through rights, tools and digital competencies and where data is pooled within common European data spaces to support research, innovation and high-quality public services [5].

A significant legal milestone in this context is the European Data Act, which came into force in January 2024. This legislation introduces a new set of rules governing access to and reuse of data generated by connected devices and digital services. Its provisions are expected to unlock significant economic value - estimated at €270 billion by 2028 - by eliminating structural barriers to data availability and fostering more dynamic data markets [5]. The Act grants consumers and businesses enhanced rights to access and control data they generate, facilitates data portability

and introduces safeguards against unauthorized cross-border data transfers. It also mandates fair pricing and promotes competition in cloud service provision, thereby lowering costs and enhancing service quality.

The projected scale of the data economy further underscores the need for a robust, EU-led infrastructure. By 2025, global data volume is expected to reach 175 zettabytes, a 530% increase from 2018 levels. In parallel, the value of the EU27 (abbreviation of European Union consisting in 27 countries, after UK left the EU on 31-01-2020) data economy is projected to grow from €301 billion in 2018 to €829 billion, with the number of data professionals rising from 5.7 million to nearly 11 million [5]. These evaluations illustrate both the opportunities and the pressures associated with digital transformation, reinforcing the need for coordinated investment, regulatory clarity and technological innovation. It is against this backdrop that the NOUS project finds both its relevance and its strategic orientation.

3.2 Overview of NOUS project

The NOUS project - a catalyst for EuropeaN ClOUd Services in the era of data spaces, high performance and edge computing - was conceived in direct response to the mounting concern over Europe's digital dependency and the strategic vulnerabilities it entails [14]. Currently, the European cloud market is overwhelmingly dominated by a few non-European hyperscalers such as Amazon Web Services, Microsoft Azure and Google Cloud. These companies collectively account for more than 75% of the European cloud service consumption, a fact that raises critical issues related to data sovereignty, economic autonomy and regulatory compliance [13]. When core infrastructures, sensitive public-sector data and industrial information are hosted by providers governed by foreign jurisdictions, the control over who accesses what data and under which conditions becomes diluted. This situation risks undermining not only European digital independence but also the capacity to enforce its own regulatory frameworks such as the General Data Protection Regulation (GDPR) and the European Data Act.

Thus, as outlined in the previous section, the NOUS project was born to satisfy the need to establish a secure, interoperable and federated cloud and data ecosystem across the continent. It proposes a new paradigm in cloud infrastructure - one that is not based on hyper-centralization and vendor lock-in, but instead on modularity, federation, openness and compliance with European values of transparency, privacy, and innovation. NOUS aims to become a cornerstone of the European cloud and data space infrastructure, capable of hosting next-generation services that require high-performance computing (HPC), edge computing capabilities and decentralized trust models [13].

3.2.1 NOUS Vision and Strategic Objectives

At its core, the NOUS project envisions a sovereign and federated cloud architecture that leverages the diverse set of computing resources already existing across Europe - including HPC clusters, quantum computing prototypes, edge nodes and IoT networks - to create a cohesive digital ecosystem. Unlike conventional cloud models, which consolidate computing and storage capabilities into a limited number of hyperscale data centers, NOUS advocates for a distributed infrastructure where computation and data storage occur closer to where data is generated. This edge-to-cloud continuum approach not only enhances performance and responsiveness, particularly for time-sensitive applications, but also supports stronger guarantees of data locality, ownership and privacy [14].

According to this vision, NOUS defines a coherent set of strategic objectives that guide its technical and societal contributions. One of the project's primary goals is to harness Europe's existing HPC infrastructure and to interface it with emerging quantum computing technologies. By unifying these powerful computational tools within a federated cloud, NOUS aspires to reinforce the European Union's capacity for scientific excellence and technological leadership in critical research areas (e.g. climate modeling, pharmaceutical innovation and sustainable material development) [13].

In parallel, NOUS dedicates substantial effort toward enabling the deployment and orchestration of Cloud-Edge-IoT technologies. This objective addresses the growing need for an infrastructure where computing and storage processes are fluidly distributed across central and peripheral nodes. The integration of edge capabilities ensures that smart devices - including sensors, vehicles and wearables - can execute computationally intensive operations locally. This architectural model not only enhances system responsiveness, but also aligns with principles of data minimization and user-centric privacy [13]. Furthermore, such distribution alleviates pressure on centralized resources and contributes to greater energy efficiency, an imperative under the EU's sustainability goals.

A third strategic axis centers on optimizing the data flow across the European digital landscape. By establishing a trusted environment that supports data interoperability, standardization and access control, the project contributes to the realization of a common European data space. This initiative underpins the broader policy objective of a unified data market in which diverse actors can collaborate, share and extract value from large-scale data assets in a legally and technically secure manner [13].

Through this tripartite strategy - advancing HPC and quantum integration, enabling edge-cloud convergence and facilitating interoperable data ecosystems - NOUS positions itself as a central enabler of the European Union's digital autonomy. Its vision is not only technological but also institutional, fostering a new generation

of cloud services that are aligned with European regulatory, ethical and operational principles.

3.3 NOUS Architectural Design

The architectural foundation of the NOUS project is conceived as an open and modular framework capable of integrating a wide spectrum of emerging technologies. It brings together cloud and edge computing, high-performance computing (HPC), quantum computing, Internet of Things (IoT), federated learning, and distributed ledger technologies (DLT) into a unified, interoperable environment. The core premise of this design is to allow the deployment of complex applications and services that can be executed at the edge - on smart devices and local computing units - while also relying on the expansive computational power provided by high-capacity back-end infrastructures. This dual capability ensures the system's responsiveness, scalability and trustworthiness across various deployment scenarios. The structure ultimately supports efficient data retrieval, processing and sharing across distributed data spaces, contributing to a dynamic and sovereign European cloud ecosystem.

Structured to be both operationally robust and methodologically flexible, the NOUS project is set to unfold over a 36-month implementation timeline. It is defined as a IaaS/PaaS/MLaaS architecture and its development methodology adheres to a two-tiered approach.

At the macro level, the project adopts a generalizable framework that can be reused by other initiatives, promoting reproducibility of best practices across the European digital ecosystem. On the micro level, the project's internal methodology is tailored specifically to NOUS objectives. It employs a modified version of the Software Development Life Cycle (SDLC), encompassing concept definition, partial implementation and validation, with the target of achieving a Technology Readiness Level (TRL) 5 [13].

NOUS's architecture is built around three core technical domains: compute, edge, and data. Each of these components plays a distinct but interdependent role in fulfilling the project's broader vision [14].

Compute component It is responsible for managing intensive computational tasks and enabling research environments. It utilizes existing HPC infrastructures while exploring integration pathways with quantum computing systems. In addition to practical elements such as authentication protocols and data transfer mechanisms, the compute layer also addresses theoretical questions related to workload optimization and system resilience. The compute infrastructure will employ multi-criteria evolutionary algorithms to dynamically allocate tasks between HPC

and quantum resources, optimizing for computational complexity, energy efficiency and infrastructural costs. Security is central to this component: investigations into darknet-based isolated networking and zero-trust security models are being pursued to ensure resilience against external threats [13].

Edge component It represents a complement that brings computation closer to where data is generated, particularly in latency-sensitive applications. It features decentralized machine learning mechanisms (federated learning and inference), designed to enable on-device training and data processing. To manage these distributed operations, the architecture leverages FastFlow, a framework for parallel programming, which provides the capacity to instantiate scalable applications using configurable parallel skeletons. The edge component also implements advanced resource orchestration strategies, which balance workload distribution between edge and cloud data centers based on system throughput, latency, and privacy constraints. Such a configuration enables the dynamic reallocation of computing tasks, preserving data sovereignty while optimizing operational efficiency [13].

Data component The data component, on which this thesis is primarily based, centers on trust, traceability and interoperability across data spaces. It integrates energy-efficient DLTs to support tamper-evident data logging, smart contract execution and real-time validation of information through oracles.

In parallel, it explores standardization mechanisms tailored to domains like mobility and energy, including the development of an auto-standardiser capable of translating heterogeneous data formats into recognized standards through recursive and heuristic algorithms. This component is architected to address both scalability and regulatory compliance. To that end, it adopts hybrid storage solutions, where blockchain is used to log hashes or metadata while the substantive data remains in off-chain repositories. This balances decentralization with the practical limitations of blockchain in terms of storage capacity and performance [13].

Mechanisms to validate data integrity - especially in environments with multiple, potentially conflicting data sources - are also included. Redundancy protocols and oracle-based verification ensure consistency particularly for critical or high-frequency data streams. In dynamic use cases, smart contracts compare newly ingested data with existing records to detect anomalies or unauthorized changes. Such capabilities underpin decentralized access control schemes, enabling secure, transparent, and policy-compliant data sharing within and across data spaces [13].

Taken together, these architectural components form a technologically and ethically grounded foundation upon which the NOUS project will deliver its vision. The combination of robust compute capacity, distributed intelligence at the edge and secure data governance mechanisms offers a powerful infrastructure blueprint for Europe's future digital ecosystem.

3.4 Use Cases

The services developed within the NOUS project are applied across multiple and heterogeneous use cases. This approach is not intended for mere illustrative purposes but serves a critical methodological function in the overall framework. Each use case is strategically selected to support the project's objectives from complementary angles, while ensuring that the technological solutions developed within NOUS are grounded in real-world conditions and diverse operational domains. More specifically, the use cases play three interrelated roles.

First, they serve as a primary source for requirements elicitation. Functional and non-functional needs associated with each NOUS technological component are derived from the practical challenges observed within these industrial and organizational contexts.

Second, they provide an empirical basis for analyzing how data lifecycle processes are currently managed in practice. This observation is crucial for informing the development of the NOUS data management architecture, especially in terms of standardization, traceability and interoperability. Through these use cases, the project captures a range of approaches to data governance, storage, transmission and reuse.

Third, the use cases act as validation environments where NOUS technologies are implemented and tested under real operational conditions. Each scenario introduces a distinct challenge, enabling the project to evaluate its technical components in complex settings. These demonstrations are not limited to theoretical proof-of-concepts but rather include applied deployments that stress-test the system's scalability, responsiveness and compliance with legal and ethical frameworks.

The selection criteria for these use cases were guided by several considerations. Participating organizations undergo digital transformation and require advanced capabilities in AI, data lifecycle analysis and federated computing. Each organization is situated in a sector aligned with one of the priority data spaces identified in the European Data Strategy, thereby linking each use case directly to the broader strategic goals of the European Union. The use cases also vary in data volume, sensitivity and complexity, offering a comprehensive testbed for the NOUS architecture.

Use Case #1 - Perception of Connected Vehicles using camera data
The first use case, which will be explored in greater depth in section 3.4.1 due to its central role in this thesis, concerns the perception of connected vehicles through distributed camera data. The goal is to improve connected vehicles perception by using real-time data collected from both roadside and in-vehicle cameras. The scenario integrates a range of NOUS components, particularly those related to edge computing and 5G-based Multi-access Edge Computing (MEC)

infrastructures. It assesses the performance of communication protocols and data orchestration between vehicles, sensors and edge servers, with special emphasis on latency, contextual AI inference and privacy-aware data handling. This use case provides a high-impact demonstration of how the edge and data components of NOUS contribute to enhancing situational awareness in mobility systems [44].

Use Case #2 - Energy Prediction and Energy Data Lifecycle Management The second use case addresses energy prediction and lifecycle management in the energy sector. The scenario involves two computationally demanding applications: forecasting short- and medium-term energy consumption and generation using weather and sensor data and managing the full data lifecycle for energy reporting. These tasks benefit from NOUS's integration with HPC for computational throughput, as well as from data standardization and blockchain-based lifecycle tracking. The use case demonstrates how NOUS can support energy providers in optimizing grid operations, reducing forecasting errors and enhancing the trustworthiness and confidentiality of energy-related data throughout its lifecycle [13].

Use Case #3 - Crisis Management and Civil Protection Platform The third use case focuses on crisis management and civil protection. It entails the enhancement of a cloud-based platform used for real-time coordination between multiple agencies during emergencies. This platform manages vast volumes of heterogeneous data, including geospatial information, live video feeds, sensor outputs and tactical communications. NOUS technologies are integrated to ensure robust access control and secure data sharing across distributed command centers and first responders. The scenario highlights the architectural flexibility of NOUS, particularly its ability to interface with legacy systems and third-party applications while maintaining data integrity and operability under high-stress conditions [13].

Use Case #4 - Scientific Data HPC Storage and AI Analytics The fourth use case is situated within the scientific research domain and targets the management and analysis of complex datasets in high-performance computing environments. It spans applications ranging from materials science to molecular simulations and nanotechnology. This use case leverages NOUS tools for data lifecycle management, federated AI analytics and HPC network. The setting involves the integration of several data types, all processed under FAIR (Findable, Accessible, Interoperable, Reusable) principles. It demonstrates how NOUS can streamline scientific workflows, reduce time-to-publication and enhance interdisciplinary collaboration across domains that are computationally intensive [13].

Together, these use cases form a multidimensional framework for evaluating NOUS. They not only span four critical sectors of the European data economy, but also address distinct aspects of NOUS architecture - computational scalability,

edge intelligence, data governance and real-world interoperability. This ensures that the platform developed within the project is both functionally comprehensive and practically viable across varied operational landscapes.

3.4.1 MASA Use Case: Edge Computing for Connected and Safe Mobility

Building on the general overview of the NOUS architecture and its four pilot scenarios, this section provides an in-depth examination of Use Case #1, developed within the Modena Automotive Smart Area (MASA). The analysis aims to show how NOUS technologies are deployed and validated in a real-world setting focused on connected mobility and road user safety.

The use case is grounded in the Modena Automotive Smart Area (MASA), a city-scale living lab in Modena, Italy, designed specifically for testing automated and connected mobility in authentic urban settings. A controllable yet realistic playground for V2X and edge-enabled services is created by MASA's instrumentation of roadway segments with digital signage, smart cameras for obstacle recognition as shown in Figure 3.1, interconnected traffic lights, and heterogeneous sensors. The location offers operational diversity and reproducibility for rigorous experimentation by combining open-street deployments, the city's Smart Model Area, with a closed test track, the Modena Autodrome's Smart Dynamic Area [45].

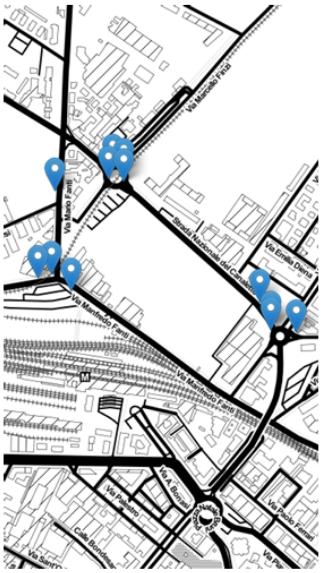

Figure 3.1: MASA's Smart Cameras [46]

Against this backdrop, the NOUS-relevant MASA use case aims to protect

Vulnerable Road Users (VRUs), such as cyclists, motorcyclists, and pedestrians, by predicting potential collisions by fusing two complementary data streams in real-time: vehicle/user geolocation (from GPS-supported client devices) and city-level detections from roadside smart cameras. In order to increase coverage to the widest range of road users and devices while lowering end-to-end latency for time-sensitive alerts, the fundamental design decision is to forego short-range, specialized V2X stacks as the sole substrate and instead employ the public 4G/5G cellular connectivity and Multi-access Edge Computing (MEC) [44].

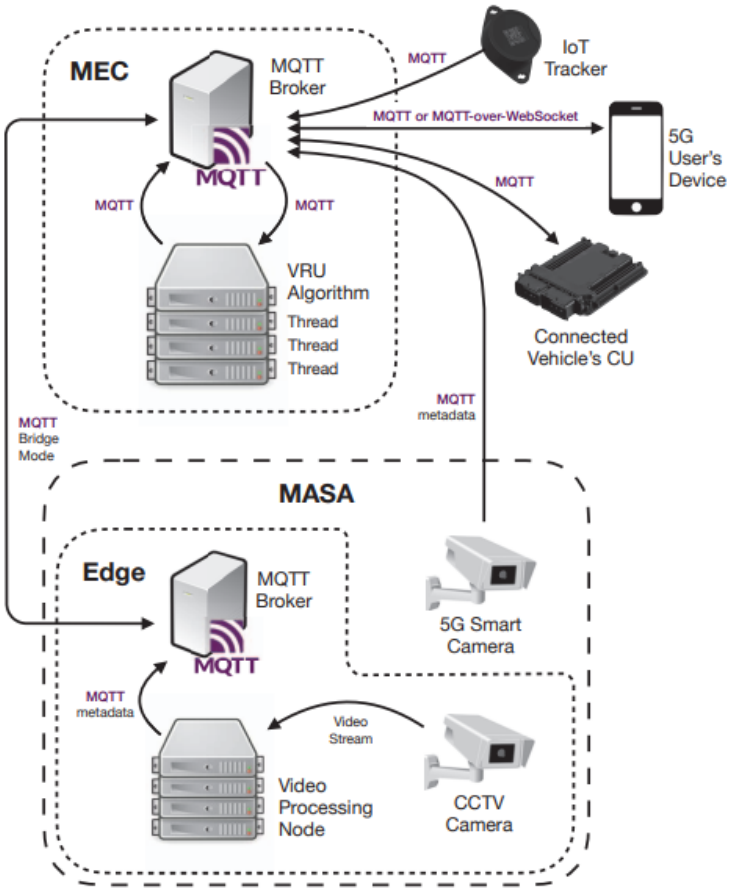

Figure 3.2: MASA's Infrastructure Architectural Design [46]

As represented in Figure 3.2, the application creates an "edge-centric" pipeline. Smart cameras in the target area perform on-device or near-device video analytics to detect and classify road users (cars, buses, motorcycles, bicycles, pedestrians), extract kinematic metadata (position, direction, speed), and publish privacy-preserving, anonymized messages to the local infrastructure. This upstream path adopts MQTT as the lightweight, publish-subscribe backbone for telemetry, allowing

the MASA infrastructure to construct a city-level mobility picture with minimal bandwidth and processing overhead. The system adopts an explicit privacy-by-design approach: only anonymous metadata are transmitted beyond the video node, aligning with GDPR constraints for public-space sensing [46].

A MEC layer, physically or topologically close to the radio access, then hosts the collision-prediction service and low-latency messaging brokers. In MASA, the MEC deployment ensures 5G coverage with high bandwidth and tight round-trip timings; in practice, MEC servers may sit several tens or even hundreds of kilometers from a base station but are fiber-connected to preserve micro-latency characteristics. Within this layer, an MQTT broker mediates mobility metadata flows, and a bidirectional WebSocket channel supports browser-based or embedded user clients that cannot natively speak MQTT, thereby covering both IoT-style devices and web apps without sacrificing responsiveness [46].

Moreover, the VRU algorithm keeps elaborating inputs from heterogeneous sources (GPS traces from participating vehicles or users and smart-camera-derived tracks) in order to forecast short-horizon collisions. Through this architecture design, per-user concurrency (one thread per connected user) and locality filtering are underlined: the usage of a spatial database allows for computations to be executed nearby road users, minimizing execution time so that updates at high frequency are feasible. Targeted alerts are emitted in case the predicted trajectories intersect within risk thresholds: the involved entities within the geographic zone receive the alerts and the loop that includes perception, prediction and notification is completed [46].

Considering now the clients' perspective, two modalities were prototyped and evaluated. First, a mobile, web-based interface, shown in Figure 3.3, was implemented and it runs in an in-vehicle tablet or smartphone. It uses GPS positions, visualizing nearby VRUs, and displaying collision warnings with a UI optimized in order to minimize the possibility of driver distraction.

Second, an embedded computing variant is connected to an addressable LED strip, yielding a less noticeable, "eyes-up" alerting surface that can complement or redundantly backstop a phone-based interface (see Figure 3.4). Both clients interoperate with the same MEC-resident services and share the same alert semantics, ensuring consistency while catering to different in-vehicle integration constraints [46].

From a networking and integration standpoint, MASA's architecture composes two broker tiers to balance locality and federation. Edge-side brokers aggregate camera-processing outputs and can bridge to the MEC broker to synchronize topics, while MEC brokers serve end-user devices over MQTT or WebSocket for the lowest achievable Round-Trip Time (RTT) on the critical alert path. This split allows multiple MEC "areas" to be added over time and lets non-latency-critical applications consume the same metadata from a central point, illustrating

Figure 3.3: MASA's Mobile Interface with Relevant Alerts and Experimental Evaluation

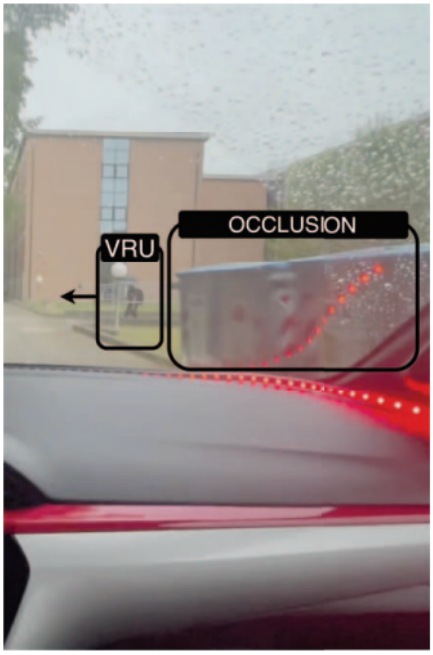

Figure 3.4: MASA's In-Vehicle Led Notifications [46]

how MASA can scale from single-junction pilots to district-wide services without redesigning the messaging substrate.

Latency is the key performance factor for safety-critical applications such as collision alerting, and the MASA experiments were designed to assess how system responsiveness changes depending on service placement and network configuration. Comparative evaluations conducted within the MASA district confirmed that deploying the alerting service at the network edge - within a MEC environment - significantly reduces end-to-end delay compared to traditional cloud-based solutions [46]. This improvement was consistently observed across different devices, communication technologies, and transmission rates, demonstrating the robustness of the edge approach for time-sensitive mobility scenarios. These findings confirm that MEC deployment represents the most effective strategy for minimizing latency in real-time safety services [46].

Complementary spatial analyses provided additional insights into network behavior under real urban conditions. Mapping latency distributions along test routes revealed that edge computing not only improves average responsiveness but also enhances its spatial uniformity, ensuring more predictable performance even in areas subject to variable network load or interference. Nonetheless, certain urban zones exhibited higher delay peaks, emphasizing the importance of coordinated optimization between edge infrastructure and the underlying communication network. Together, these results highlight the central role of edge computation in achieving reliable and low-latency service delivery for connected mobility [46]. Overall, the analysis highlights that while edge computing substantially improves average responsiveness, residual latency peaks may depend on transient network conditions rather than architectural design.

A crucial aspect that emerges is the robust architectural design through which MASA empowers the trustworthy city sensing. Namely, video analytics are kept at the edge, and only metadata proceed into the messaging layer; hence, identity is decoupled from motion tracks and message formats and brokers are industry-standard and horizontally scalable. This approach is aligned with GDPR principles for conscious smart-city design and is directly transferable to other NOUS pilots where sensitive signals must be processed close to the source.

The use case also validates the practicality of cellular-first V2X safety services. Public 4G/5G networks combined with MEC can reach a far larger set of participants than dedicated short-range stacks alone, including pedestrians and cyclists carrying only commodity smartphones. By moving compute nearer to the radio and federating MASA's smart cameras through MQTT, the system achieves alert latencies that are consistent with driver reaction-support time budgets, while retaining the ability to interoperate with more specialized in-vehicle units when present [46]. This hybridizes the best of both worlds and aligns with the NOUS vision of an edge-cloud continuum serving heterogeneous data producers and consumers.

Crucially, MASA as a facility underwrites these outcomes with its layered infrastructure. The Smart Model Area's city streets provide everyday complexity - occlusions by buses, variable pedestrian flows, signalized intersections - while the Smart Dynamic Area allows controlled reproduction of edge cases and repeatable benchmarking [45]. The emerging data center and Smart City control room provide local hosting and integration with municipal services, enabling functional tests that consider not just raw latency, but also maintainability and governance factors that real cities must manage. This combination - open roads, closed track, local compute - gives MASA a unique role.

Finally, the MASA use case demonstrates a complete blueprint for NOUS pilots that need to turn heterogeneous and sensitive data stream into real-time safety value. It shows how the environment can be integrated with edge perception, keeping only anonymous, standardized metadata over lightweight brokers. Additionally, this use case addresses the latency factor by leveraging prediction services at MEC nodes in order to, lastly, deliver alerts through multimodal clients that fit different vehicle contexts. This blueprint is already functioning in Modena's living lab, and its empirical results justify its replication as NOUS scales sovereign cloud-edge services within European data spaces.

This chapter has described the strategic, architectural, and operational framework of the NOUS project, highlighting how it leverages cloud, edge, and data technologies in the context of a common vision of sovereignty and interoperability. Within this circumstances, the NOUS *data component* plays a pivotal role by integrating blockchain mechanisms for secure and transparent data lifecycle management.

The Proof of Concept presented in the chapter 4 is designed as part of this component: it operationalizes NOUS's principles by testing a blockchain-enabled data flow that connects data sources - in this case elaborated in the MASA use case - with federated dataspace services. Through this implementation, the research moves from the architectural foundations of NOUS to their concrete realization in a prototype environment.

Chapter 4

A Blockchain Framework for Data Lifecycle Management in Data Spaces

This chapter presents the architectural and implementation framework developed to support data lifecycle management within federated data spaces, integrating blockchain-based mechanisms for the notarization and traceability of operations. A top-down approach is adopted, beginning with the overall architectural vision and progressively delving into increasing levels of technical detail. Starting from the system architecture and reference frameworks - EBSI and Simpl-Open - the chapter gradually transitions towards the modelling of operational processes, the incremental validation framework, and finally, the detailed implementation that reproduce the complete end-to-end execution process.

To ensure terminological consistency, a Glossary is provided in the Appendix A.2, collecting the technical terms introduced throughout this chapter.

4.1 System Architecture

This section illustrates the conceptual and technical foundations that guided the design of a reference architecture capable of sustaining a trustworthy and auditable data lifecycle within complex, multi-domain environments, by combining the potential of data spaces with the trust and traceability features of blockchain technologies. The model builds upon a federated and modular approach, combining governance, interoperability, and distributed trust mechanisms.

The architectural framework follows three guiding principles:

- Separation between control and data layers, ensuring scalability and modularity.
- Integration of blockchain-based trust services - for notarization, provenance, and verifiability - without introducing the complexity of tokenization or digital wallets for end users.
- Use of a shared, semantically interoperable data space, which provides the foundation for cross-domain data exchange and governance enforcement through the NOUS ecosystem.

The overall infrastructure, depicted in Figure 4.1, represents a federated and scalable environment that guarantees the transparency, auditability, and long-term sustainability of data operations. It integrates five interdependent layers: governance entity, distributed ledger, orchestration middleware (data space), standard APIs, and domain-specific use cases.

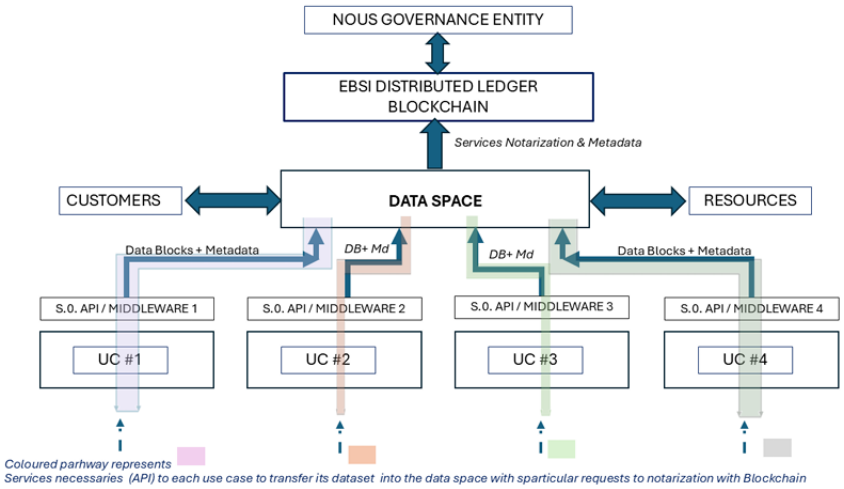

Figure 4.1: System Architecture

Governance Layer

At the top of the architecture lies the Nous Governance Entity, which serves as both a policy and trust orchestrator. It ensures alignment with European Data Strategy objectives, supervises the roles of stakeholders, and enforces operational compliance. Beyond its normative function, it acts as a technical anchor for federation management, maintaining a shared trust fabric across distributed participants. In practical terms, this entity defines and enforces the contractual and semantic

rules that regulate participation, enabling transparent accountability mechanisms across all connected domains.

Distributed Ledger Layer

Below the governance layer operates the EBSI Distributed Ledger Blockchain, detailed in Section 4.1.1 which provides the core trust and traceability services. It guarantees notarization, timestamping, and verifiability of transactions, along with immutable metadata storage. Through cryptographic validation, each interaction between data providers and consumers becomes independently auditable and tamper-proof, supporting regulatory principles of data sovereignty and accountability.

EBSI - the European Blockchain Services Infrastructure - was selected as the underlying ledger technology because it ensures alignment with EU digital policies, native interoperability with public services, and official endorsement as the European reference infrastructure for blockchain-based trust services. This choice also allows seamless integration with existing eIDAS (Electronic Identification, Authentication and Trust Services) and verifiable credentials frameworks, creating the basis for cross-border trust [47].

Orchestration Layer: The Data Space

The central Data Space layer acts as a semantic and technical broker, harmonizing access to data and managing interoperability across organizational boundaries. It communicates with the blockchain to record metadata, invoke notarization services, and enforce governance rules. This layer abstracts the underlying data infrastructure, allowing organizations to share and consume data securely while retaining full control over their assets.

The architecture remains agnostic with respect to the specific data space technology, making it adaptable to different frameworks or commercial platforms.

Middleware and API Layer

Each Use Case (UC) developed within the NOUS ecosystem - and designed to be extensible to other scenarios, connects to the data space through a dedicated middleware/API interface, which acts as a translation and compliance layer. These components integrate heterogeneous local systems - such as IoT networks, digital twins, or enterprise repositories - by encapsulating their data into structured blocks enriched with metadata. They also validate conformance to governance rules and trigger blockchain notarization through the data space layer.

In the reference diagram, four UCs are depicted in alignment with the NOUS project environment, though the design supports dynamic onboarding of new use

cases without any architectural redesign, enabling continuous scalability.

Providers and Consumers

On the system's edges, Providers supply data or expose services through their middleware layers, while Consumers access them via the data space, performing queries or subscribing to specific data streams. All interactions occur under governance enforcement, and every dataset is accompanied by provenance metadata and, when relevant, by smart contract terms defining usage conditions. This ensures transparent, traceable data exchanges aligned with the FAIR principles (Findable, Accessible, Interoperable, Reusable).

Federated Design and Operational Principles

From an operational standpoint, the architecture supports federated deployment: each node - be it a provider, consumer, or use case - retains autonomy over its infrastructure, yet adheres to common interoperability and trust protocols. This design achieves a balance between centralized coordination (via the data space and governance entity) and decentralized validation (through blockchain), ensuring both efficiency and independence.

The model also enables off-chain storage of sensitive data, with only metadata and hashes committed to the blockchain. This approach guarantees compliance with privacy principles such as data minimization, while maintaining verifiability and auditability through cryptographic proofs.

Beyond basic operations (create, update, delete), the architecture embeds advanced mechanisms for data lineage and lifecycle tracing through blockchain integration. Each transformation performed within the system - including AI-based processes - is registered on the distributed ledger, linking derived datasets to their original sources.

This creates a complete and immutable chain of provenance, allowing the reconstruction of every processing step a dataset undergoes, from ingestion to transformation and eventual deletion. Such end-to-end traceability is fundamental not only for technical transparency but also for regulatory compliance with frameworks like the GDPR and the upcoming AI Act.

By preserving these verifiable links, the system fosters trustworthy, auditable, and explainable AI-driven services, capable of operating reliably even in highly distributed and multi-actor environments.

4.1.1 The European Blockchain Services Infrastructure - EBSI

The European Blockchain Services Infrastructure (EBSI) is a flagship initiative of the European Commission, jointly developed with the EU Member States, Norway, and Liechtenstein, to establish a cross-border, trustworthy blockchain infrastructure for public services in Europe. EBSI aims to enhance the efficiency, transparency, and trustworthiness of digital transactions between citizens, businesses, and public administrations, leveraging blockchain as a tamper-evident, decentralised trust layer.

Rather than focusing on a single application, EBSI represents a federated network of blockchain nodes operated by public institutions, each acting as a participant in a shared European infrastructure. This network underpins a series of public-service use cases - from digital identity management to credential verification and document traceability - ensuring that individuals and organisations can interact securely while maintaining control over their data [47].

A key design principle of EBSI is its adherence to open standards and self-sovereign identity (SSI) principles. Through a verifiable credentials (VC) framework, users can hold and present trusted digital attestations using digital wallets, while verifiers can instantly validate the authenticity and revocation status of those credentials via the blockchain. The architecture ensures that personal data remains off-chain, aligning with European data protection regulations such as the GDPR, while the blockchain itself serves as an immutable proof layer supporting trust and auditability [47].

In this way, EBSI represents not only a technological infrastructure, but a policy instrument that operationalises the European values of transparency, sovereignty, and privacy in the digital domain.

Architecture and Core Components

Technically, EBSI is implemented as a public-permissioned blockchain network, meaning that write permissions are restricted to authorised nodes operated by public entities, while data transparency and auditability are guaranteed through open read access. This model achieves a balance between accountability and decentralisation, ensuring institutional trust without compromising openness.

Each participating Member State operates one or more national nodes, coordinated by an EU master node, forming a federated infrastructure capable of secure cross-border operations. These nodes host the core services of the network, including transaction validation, registry services, and credential status management (such as revocation registries).

At the application layer, EBSI implements a standards-based identity and credential framework, relying on:

- Decentralized Identifiers (DIDs) and Verifiable Credentials (VCs) as defined by the W3C for decentralised identity management;
- OpenID for Verifiable Credential Issuance (OID4VCI) and OpenID for Verifiable Presentations (OID4VP) for interoperability across wallets and issuers;
- Digital wallets that allow citizens, students, and organisations to securely store and control their credentials.

This standards stack ensures interoperability with other European initiatives and avoids dependency on proprietary technologies, allowing EBSI to act as a neutral, open trust layer upon which other infrastructures—such as Simpl-Open or sectoral data spaces—can build.

EBSI's blockchain layer provides the proof-of-existence and proof-of-trust functions: credential issuers register their identities on-chain, while verifiers use these records to confirm authenticity and revocation status. Importantly, no personal data is stored on-chain, preserving compliance with European privacy requirements while ensuring full verifiability of transactions and credentials.

Within the broader ecosystem described in this thesis, EBSI complements the middleware layer by providing the trusted credential and identity layer that supports federated data interoperability.

In practical terms, this means that EBSI could be integrated within federated data infrastructures to verify the authenticity and authorisation of participants and to ensure that all data-sharing operations are anchored in a secure and verifiable trust framework.

Consequently, in the context of this thesis, EBSI represents the trust and credential verification layer within the federated architecture model. It provides the technical and regulatory foundation for secure cross-border cooperation among European data spaces, making it a cornerstone of the European vision for a trusted and interoperable digital ecosystem.

4.1.2 Simpl-Open: A Federated Middleware Framework for European Data Interoperability

Simpl-Open is a large-scale, open-source middleware framework funded by the European Commission with the objective of enabling interoperability and federation across European data spaces. Rather than representing a single dataspace, Simpl-Open acts as a federated infrastructure layer - a socio-technical framework designed to connect autonomous data ecosystems through shared standards, governance models, and open-source components.

The initiative contributes to Europe's long-term strategy for digital sovereignty and data autonomy, serving as the technical backbone of the envisioned European

Cloud Federation. Simpl-Open provides the mechanisms for secure, efficient, and trusted cross-domain data exchange between public, private, and research actors. Its ultimate goal is to federate data, applications, and infrastructures while maintaining ownership and control at the source, thereby fostering trust and resource efficiency within the European digital ecosystem.

Architectural Vision

Simpl-Open is conceived as a modular, multi-vendor, and interoperable architecture that enables the design, deployment, and interconnection of European data spaces. It functions as a middleware infrastructure that sits between individual data spaces and the underlying cloud-to-edge computing resources.

At the architectural level, Simpl-Open is structured around five layers, each addressing a distinct set of capabilities:

- Integration Layer: provides the foundational capabilities for secure and trusted interaction among participants, including discovery, access control, and federation management.
- Data Layer: enables the exchange, sharing, and management of data resources and applications, ensuring semantic and syntactic interoperability between providers and consumers.
- Infrastructure Layer: connects to external computing and storage resources without hosting them directly, allowing participants to execute workloads on distributed environments.
- Administration Layer: supports the operation of the other layers through monitoring, configuration, and orchestration tools.
- Governance Layer: defines cross-cutting mechanisms for compliance, coordination, and resilience, including human-in-the-loop functions.

Through this layered model, Simpl-Open provides the shared middleware services and policy framework upon which independent data spaces can be built and interconnected. Each data space can extend Simpl-Open with domain-specific components - such as semantic vocabularies, certification mechanisms, or data-quality assessment services - while preserving overall interoperability.

Actors and the Simpl-Open Agent

To operationalize this federated vision, Simpl-Open defines five main actor groups, shown in Figure 4.2, that collectively represent a distributed network of cooperating

entities within an open ecosystem. Each actor interacts through a local middleware component called the *Simpl-Open Agent*, which serves as the gateway to the data space.

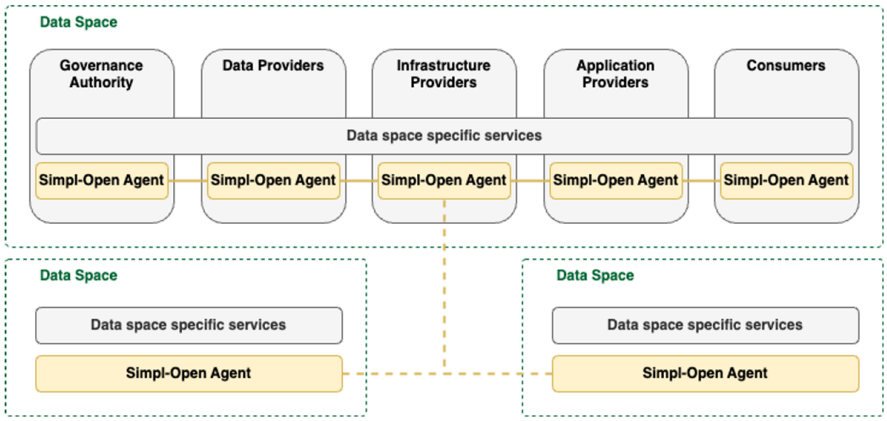

Figure 4.2: Simpl-Open Architecture Vision: Actor Groups and Cross-Dataspace Interoperability

The five actors considered by Simpl-Open are:

Governance Authority: the entity responsible for defining the rules, standards, and procedures for participation within a data space. It validates the identity of participants and ensures compliance with governance policies.

Data Providers: organizations or entities that supply data assets to the ecosystem, making them available to other participants under controlled access and usage policies.

Infrastructure Providers: actors offering computational or storage resources (e.g., cloud or edge infrastructure) that other participants can access via Simpl-Open interfaces.

Application Providers: entities that develop or deliver applications and services consuming or enriching data within the data space.

Consumers: the final users or organizations that access, process, or reuse data and applications provided by others in the ecosystem.

Each of these actors operates one or more nodes - that is, distinct sets of IT resources that can be geographically distributed - and deploys a Simpl-Open Agent locally. The Simpl-Open Agent acts as the middleware connector responsible for secure communication, authentication, and authorization within the data space. It allows the participant to discover services, exchange data, and apply access policies defined by the governance authority.

Identification and trust are central to this architecture. The Governance Authority validates the digital identity of each participant and issues a verifiable

credential that the participant installs in its Simpl-Open Agent. This credential enables authentication toward other members and the enforcement of authorization rules for data access and service usage.

Simpl-Open remains agnostic to the specific implementation of a given data space. The *data-space-specific services* shown in the figure 4.2 represent customizable layers that can define data representation standards, quality certifications, or peer-review rules, adding value beyond the simple exchange of assets. Multiple data spaces integrating Simpl-Open can also interconnect with each other, allowing cross-dataspace interoperability, where services and assets traverse the boundaries of individual ecosystems under common governance.

Architectural Principles and Design Rationale

Simpl-Open's design is guided by ten architectural principles that ensure openness, resilience, and sustainable interoperability across heterogeneous domains. These principles can be summarized as follows:

- Federation: Simpl-Open connects autonomous yet interoperable entities under shared standards and legal frameworks, preserving participant autonomy while enabling collaboration.
- Modularity and Loose Coupling: the system is built from independent, replaceable components linked via standardized APIs, ensuring adaptability and minimal dependency.
- Openness and Technological Agnosticism: specifications are open and technology-neutral, allowing deployment in diverse environments and preventing vendor lock-in.
- Composability and Extensibility: services can be combined or extended, supporting iterative evolution and community contributions to the framework.
- Interoperability and Discoverability: common interfaces and registries make all services discoverable and reusable across different data spaces.
- Scalability and Resilience: the system supports both vertical and horizontal scaling while maintaining fault tolerance and high availability.
- Security, Privacy, and Trust: built-in mechanisms ensure confidentiality, integrity, and compliance with European regulations, fostering trust among participants.

Together, these principles position Simpl-Open as a federated, open, and secure middleware framework, capable of sustaining Europe's interconnected data infrastructure in the long term.

Position within the European Data Ecosystem

Simpl-Open should not be understood as a data space itself, but rather as the interoperability infrastructure that enables the creation and interconnection of multiple data spaces. It provides the shared middleware, governance rules, and interfaces necessary for cross-domain cooperation.

In this regard, Simpl-Open functions as the conceptual and architectural reference model, whereas specific technologies - such as Sovity's Eclipse Dataspace Connector (EDC) presented in Section 4.3.1 - constitute practical implementations that operationalize its principles. Therefore, while Simpl-Open provides the theoretical framework depicted in the diagrams included in the Appendix and explained in Section 4.1.3, the Proof of Concept developed for this thesis employs Sovity's platform to instantiate those concepts within an experimental environment. This alignment between Simpl-Open's conceptual architecture and Sovity's technical realization exemplifies the European approach to federated design - where common frameworks and open standards enable diverse implementations to interoperate within a cohesive, trusted ecosystem.

4.1.3 Processes Design

The operational design developed within this thesis focuses on a selected set of processes that were modelled according to the Simpl-Open framework, which defines the reference architecture and operational logic for European federated data spaces. Simpl-Open specifies a variety of participant and resource management operations aimed at ensuring interoperability, trust, and accountability within data-sharing ecosystems.

In the context of this work, four key operations were analysed, modelled, and adapted in the diagrams depicted in figures in the Appendix A.3:

- Participant Onboarding
- Resource Creation
- Resource Update
- Resource Withdrawal

These processes were selected as they constitute the core operational lifecycle governing the interaction between data providers and the underlying data space

infrastructure. Rather than reproducing Simpl-Open's workflow logic, this work adopts its operational framework as a conceptual and procedural foundation from which to derive an extended, blockchain-enhanced model.

The methodological approach consisted of a detailed process-level analysis of Simpl-Open's specifications, with the objective of identifying the functional intersection points where the inherent characteristics of blockchain technology - immutability, integrity assurance, and public verifiability - could meaningfully complement the framework's existing trust mechanisms. In other words, each Simpl-Open process was decomposed into its constituent phases, and the stages involving critical state transitions (such as participant validation, resource publication, update, or revocation) were isolated as suitable integration points for a notarization action.

By embedding a notarization layer at these precise points, the Simpl-Open conceptual model was extended into a hybrid operational architecture. This new model preserves the federated design and interoperability principles of Simpl-Open, while augmenting them with distributed ledger capabilities that enable tamper-evident, auditable event tracking across the federation.

The resulting architecture is articulated along three interdependent layers:

- Provider, which generates the operational event and in the specific case of this thesis is identified as MASA Use Case, described in Section 3.4.1;
- Data Space Layer, which manages and governs the transaction according to Simpl-Open's logic;
- Blockchain Layer, which notarizes the event by recording its cryptographic hash and metadata on-chain.

Through this layered structure, every relevant operation acquires an additional dimension of verifiable accountability. The blockchain acts as a trust anchor external to the data space, ensuring that all recorded events are immutable, independently auditable, and cryptographically verifiable without disclosing sensitive data.

This approach effectively transforms Simpl-Open's federated middleware into a verifiable trust framework, where blockchain technology strengthens governance transparency and operational integrity. The resulting system architecture, illustrated in the previous section, therefore represents a pragmatic extension of the Simpl-Open model, bridging the domains of federated data interoperability and decentralized trust assurance.

Onboarding Process

The onboarding process in Simpl-Open governs how a new entity - whether a data provider or consumer - becomes an authorised participant within a data space.

The procedure begins with the submission of an onboarding request to the Governance Authority, which acts as the trusted root of the data space. This request includes identity details and relevant metadata describing the participant's role and resources. The Governance Authority is responsible for identity verification, credential issuance, and authorization setup, ensuring that only validated participants gain access to the ecosystem.

Once verification is completed, the participant's Simpl-Open Agent is configured and registered within the federation. The agent receives digital credentials or tokens that allow it to authenticate securely and interact with other components of the data space. This step establishes the trust relationship between the participant and the governance framework, making onboarding a critical entry point for maintaining security, accountability, and interoperability.

Add a Resource

The Add-a-Resource process in Simpl-Open defines how a data provider publishes a new resource, typically a dataset, into the dataspace catalogue, ensuring its compliance with governance and interoperability standards.

The process starts when a Provider, such as MASA, prepares a Resource Description using the official resource schema defined by the Governance Authority. This schema outlines the required and optional metadata fields that guarantee uniformity and discoverability within the federation. It specifies elements such as identifiers, dataset title, format, distribution endpoint, provenance, and links to related datasets. Providers can also include additional metadata like intended users, access or usage policies, and Usage Contract templates to define the legal and operational conditions under which the data can be accessed.

Once the Resource Description is complete, the Provider digitally signs and submits it to the Governance Authority. The Authority performs a concise but comprehensive validation process covering syntax, semantics, and quality. The syntactic check ensures that fields follow the correct format and structure; the semantic check verifies consistency with reference vocabularies and governance models; and the quality check confirms completeness and compliance with defined standards.

After successful validation, the Resource Description is published on the Dataspace Catalogue, making it discoverable and usable by other authorized participants. The Provider then receives notification of publication, confirming that the resource has been officially registered and is available under its declared usage conditions.

This process ensures that every published asset follows a common metadata model, strengthening interoperability, transparency, and trust across the Simpl-Open ecosystem.

Add a New Version of a Resource

The Add-New-Version process in Simpl-Open defines how a data provider updates an existing resource while preserving the integrity and immutability of its previous versions. Since every Resource Description carries a unique digital signature generated at the time of its creation, it cannot be altered after publication. This principle guarantees traceability and prevents any unauthorised modification of registered assets.

Accordingly, an update does not overwrite the existing resource but creates a new version with a new immutable signature. The former version is deprecated within the dataspace registry, while on the blockchain the previous transaction remains permanently recorded. The new transaction, representing the updated version, is added as a new Add operation and cryptographically linked to the hash of the preceding version, thus maintaining an auditable and verifiable history of the resource lifecycle.

The process begins when the Provider retrieves the current Resource Description and the latest version of its schema from the registry. The schema serves as the structural template for the new version, ensuring that the updated description complies with the most recent data model and governance standards. Based on this schema, the Provider modifies the content as required, updating metadata, refining quality attributes, or revising descriptive details. Resource-specific elements such as data provenance, distribution format, access policies, or usage contracts can also be adjusted to reflect the resource's evolution.

Once the new description is ready, the Provider digitally signs and submits it to the Governance Authority for validation. The validation follows the same structured sequence used for new resources but in a more targeted manner, thus syntax, semantics and quality are verified.

When validation is successfully completed, the Governance Authority deprecates the previous version of the resource in the registry and publishes the new one in the Dataspace Catalogue. The new version immediately replaces the old one in discovery and access services, while all previous versions remain accessible for audit and traceability purposes. Finally, the Provider is notified of the successful update, confirming the publication and activation of the new version.

This process ensures a controlled and transparent versioning mechanism within Simpl-Open: resources evolve through new immutable records rather than modification of existing ones, maintaining both the continuity of information and the integrity of the federated dataspace.

Revoke Access to a Resource

The Revoke-access process in Simpl-Open defines how a data provider removes a published resource from the dataspace while maintaining full traceability and

compliance with blockchain immutability principles. In the dataspace, withdrawal corresponds to the removal of the resource description from the catalogue, ensuring that it is no longer visible or accessible to other participants.

However, since the blockchain ledger is immutable, a resource cannot be deleted once its hash and metadata have been notarized. To address this, the system implements a Revoke transaction, which serves to formally invalidate the resource and block any further access to it. The revoke transaction is recorded as a new block that is cryptographically linked to the previous one containing the original resource hash, thereby preserving the integrity and continuity of the historical record. The process begins when the Provider submits a revocation request to the Governance Authority, indicating that a specific resource must be withdrawn from publication. Following this request, the Governance Authority executes the withdrawal operation within the dataspace, effectively removing the resource description from the active catalogue. This ensures that the resource can no longer be discovered or accessed by other participants through standard data sharing mechanisms. At the same time, the Revoke transaction is registered on the blockchain, permanently recording the event and linking it to the prior record of the resource. This guarantees transparency and non-repudiation of the withdrawal action, while upholding the immutability of the audit trail. Once the withdrawal and blockchain revocation are completed, the Provider is notified of the successful operation. The process thus ensures that obsolete or invalid resources can be securely and transparently decommissioned, maintaining both operational control in the dataspace and verifiable accountability on-chain.

The described operational processes collectively define the functional backbone of the Simpl-Open framework. Together, they establish a controlled lifecycle for resources, from participant enrolment and data publication to version management and revokation. Each process is designed to ensure transparency, accountability, and interoperability between the governance framework and technical infrastructure. This structured foundation provides the procedural continuity upon which the Proof of Concept implementation builds, translating the conceptual design into an operational workflow supported by blockchain-based notarization and federated governance mechanisms.

4.2 Proof of Concept (PoC) Implementation Steps

The Proof of Concept (PoC) developed within the NOUS project represents a foundational milestone in demonstrating the feasibility of embedding blockchain-based notarization within federated data spaces. Although it constitutes an early-stage prototype, this Minimum Viable Product (MVP) already holds significant

innovative potential, as it introduces a conceptual and operational framework that has not been previously proposed in the literature or in practice.

By implementing verifiable lifecycle management through a blockchain-enabled trust layer, the PoC demonstrates how compliance, accountability, and data sovereignty can be architecturally enforced within federated environments. As such, it stands not merely as a preliminary implementation, but as a proof of innovation that validates the architectural vision underpinning the broader NOUS initiative.

This section presents the incremental validation and evolution strategy for the PoC, which is structured around two tightly interrelated dimensions of development:

- **Blockchain validation axis** - focusing on the progressive consolidation of the on-chain lifecycle management framework, achieving increasing levels of technological and governance maturity (see Section 4.2.1).
- **Interaction layer evolution axis** - addressing the transformation of the communication interface between the data space middleware and the blockchain layer, which directly influences scalability, resilience, and auditability (see Section 4.2.2).

While the blockchain validation dimension focuses on the incremental maturation of the trust and notarization layer, the interaction layer dimension concerns the architectural evolution of the interoperability mechanisms enabling reliable communication between data space and blockchain components. Together, these two interrelated axes define the evolutionary trajectory of the PoC toward a scalable, auditable, and governance-compliant architecture.

The progressive evolution presented here follows a phased logic inspired by the Technology Readiness Levels (TRLs) methodology, adopted in European innovation projects to structure technological consolidation through measurable milestones. Functional correctness, distributed interoperability, and institutional alignment are three aspects that are improved by each phase, guaranteeing that the blockchain and the communication subsystems develop within a unified architectural framework. This implementation represents an incremental strategy based on iterative validation, evidence-based evolution, and continuous risk mitigation across technical and governance layers.

4.2.1 Blockchain Incremental Validation Framework

The incremental validation framework governs the progressive maturation of the blockchain component, from local deterministic simulations to deployment within institutional, permissioned environments (see Figure 4.3). The objective is to demonstrate that data lifecycle events within the data space - such as creation, update, and withdrawal of resources - can be reliably notarized through smart

contracts operating in compliance with European trust and regulatory frameworks.

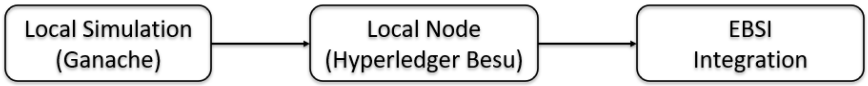

Figure 4.3: Blockchain Incremental Phases

Ganache, a popular development tool that simulates blockchain behaviour in a local sandbox, is used to perform the functional validation at this stage of TRL 4 in a controlled, Ethereum Virtual Machine (EVM) compatible local environment. Because this configuration is technically in line with the EVM standard, any smart contract that is implemented and tested in this setting is going to be compatible with actual blockchain networks like Hyperledger Besu or EBSI and will subsequently be deployed on them.

The use of an EVM-compatible environment such as Ganache is particularly valuable in the early phase for several reasons: (i) it enables deterministic and repeatable testing, since block generation and transaction ordering can be fully controlled; (ii) it provides instant feedback on contract execution, event emission, and gas consumption; (iii) it allows developers to validate the logical integrity of smart contracts before incurring the computational overhead and governance constraints of live networks.

In this controlled environment, each data lifecycle operation is executed as a blockchain transaction that emits structured events containing hashed identifiers and associated metadata. The validation process focuses on verifying:

- Semantic accuracy and correctness of smart contract functions.
- Event consistency and integrity of state transitions.
- Proper linkage between dataset versions through cryptographic chaining.

In this phase, the smart contract's validated baseline model is established, ensuring that it satisfies the lifecycle protocol essentials and provides a solid basis for the upcoming distributed testing.

TRL 5, the second validation phase, moves closer to testing a permissioned network with Besu. It extends the PoC from local deterministic execution to a real blockchain node deployment using Hyperledger Besu. This setting ensures complete adherence to real-world circumstances, network consensus mechanisms, and transaction persistence while deploying and executing the smart contract inside an actual blockchain network.

During this stage, a local Besu test node is created and it acts as an intermediate bridge environment before full interoperability with the EBSI testnet.

There are several advantages using Besu as an intermediate validation step:

- It lets testing under realistic consensus conditions, simulating permissioned governance.
- It verifies transaction propagation, block confirmation latency, and event finality.
- It validates identity management and access control mechanisms aligned with permissioned and governance-driven blockchain infrastructures;
- It establishes a technical bridge between local prototype testing and institutional permissioned frameworks (EBSI).

The functional metrics of transaction latency, gas efficiency, and event propagation time are evaluated at this level. Additionally, the reliability of middleware connectors' interactions with the permissioned ledger via standardised APIs are verified. As a result, NOUS participants can deploy the distributed and interoperable configuration in cooperative governance scenarios.

Reaching a TRL 6 of system demonstration in a relevant environment, the final stage targets full integration with the EBSI environment, thereby validating operational viability within a European trust and governance context. Beyond technical interoperability, this phase assesses compliance with institutional frameworks, including:

- Node governance models and operational policies.
- Decentralized Identifiers (DIDs) and Verifiable Credentials (VCs).
- Adherence to principles derived from the GDPR and the DGA.

Validation at TRL 6 confirms that the blockchain-based lifecycle management system can operate within a federated trust ecosystem, meaning that multiple independent entities share infrastructure while preserving accountability and transparency.

This incremental approach ensures that the PoC progresses through risk-controlled, evidence-based maturity stages, coordinating technological advancement with compliance and governance readiness.

4.2.2 Evolution of the Interaction Layer

The interaction layer, which establishes the communication mechanism between the blockchain layer and the data space middleware, is the subject of the second evolutionary axis. Scalability, fault tolerance, and auditability are all directly impacted by this interface's design. For a concurrent development, the interaction layer must change from simple synchronous communication to asynchronous, message-oriented interoperability, as the blockchain component matures through incremental TRLs (see Figure 4.4).

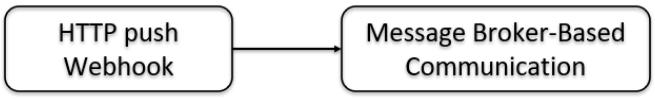

Figure 4.4: Interaction Layer Incremental Phases

Webhook-Based Integration

In the early configuration, the data space communicates with the blockchain through API calls handled by a webhook service. Each time a lifecycle operation occurs the middleware sends an HTTP request to the webhook endpoint. The webhook component acts as an intermediary between the data space layer and the blockchain, performing the following core functions:

- Intercept and log every API request and response exchanged with the Dataspace Portal.
- Store structured metadata for auditing, debugging, and analytics.
- Automatically trigger a smart contract on the blockchain when specific API events occur.
- Guarantee the immutability and verifiability of transaction records through blockchain anchoring.

This mechanism allows immediate synchronization between data space operations and blockchain notarization. Its deterministic behavior and minimal integration overhead make it particularly suitable for early-stage component validation (TRL 4), facilitating debugging and end-to-end verification of event semantics. However, webhook-based communication exhibits inherent limitations when applied to distributed and federated contexts:

- Network dependency and reliability limitations: the webhook mechanism relies on stable, synchronous network connections; temporary outages or latency spikes can interrupt the transaction flow, causing missed or duplicated events.

- Lack of persistence and delivery assurance: webhook calls are executed in real time without reliable buffering or delivery guarantees, making it impossible to replay or recover events in the event of a failure.
- Limited control over ordering and idempotency: events might be retried or received out of order, which could lead to redundant blockchain transactions and compromised chronological consistency.

The shift to a more resilient, asynchronous model is encouraged by these flaws, which limit scalability and resilience.

Transition to Message-Oriented Middleware

The next phase of development plans the introduction of a message-driven communication substrate, implemented through a messaging system (such as Apache Kafka), in order to overcome the drawbacks of synchronous webhooks. This setup allows for the reliable, asynchronous propagation of lifecycle events through persistent topics between data space connectors and blockchain adapters.

The integration flow unfolds as follows:

- API response publication: when the backend receives a response from an external API or process, it publishes the response as a message to a Kafka topic rather than directly returning the result as an HTTP response.
- Kafka consumer service: a dedicated consumer service subscribes to the topic and listens for new messages. Upon message arrival, it parses the payload, performs validation, and applies any required data transformation or enrichment.
- Smart contract triggering: once validated, the appropriate smart contract is triggered through the blockchain integration module (e.g., Web3 interface or an on-chain SDK). This process is fully asynchronous, ensuring decoupling between API handling and blockchain execution.
- Error handling and reliability: Kafka's inherent reliability mechanisms guarantee message delivery semantics. Failed transactions or processing errors are redirected to a dead-letter topic for monitoring, analytics, and potential reprocessing.

The key benefits that the transition to the message-based solution entails are:

- Decoupled Architecture: the backend is no longer dependent on synchronous webhook calls.

- Scalability: Kafka enables distributed and horizontally scalable event processing.
- Resilience: temporary failures in API or blockchain layers do not block message processing.
- Replay Capability: historical events can be replayed for auditing, debugging, or data recovery.

This transition enables a reliable and auditable event flow, maintaining compliance with minimal-logging principles while strengthening observability and operational robustness. Moreover, the message-oriented approach aligns with EBSI's communication and interoperability guidelines, paving the way for future integration with institutional message buses and verifiable credential exchange mechanisms.

This evolution marks the shift from controlled testing environments (TRL 4-5) to pre-deployment readiness (TRL 6+), being a significant step in the incremental roadmap. It formalizes the technical and operational maturity of the PoC's communication architecture and it ensures that requirements such as scalability, resilience, and auditability are fully met. This represents a crucial aspect since these conditions are crucial for large-scale and multi-node deployments, both across NOUS data spaces and in future European federated infrastructures. Overall, the shift from webhook-based to message-oriented solution embodies not only a technical enhancement but also a strategic advancement toward production-level interoperability and governance alignment.

4.2.3 Data Provider Integration Layer

Within the overall architecture, and in alignment with the Simpl-Open framework, an additional component is foreseen to enable seamless interaction between the MASA data provider and the NOUS data space. This module, acting as a connector, constitutes a critical enabler for integrating external data sources into the federated ecosystem and for bridging the operational gap between data generation and notarization.

The connector is designed to act as an interoperability gateway between MASA's message-oriented environment and the data space. Its primary function is to capture events produced within the MASA infrastructure and translate them into standardized API operations recognizable by the dataspace middleware.

The MASA Connector operates as an external component within the architecture, managing two complementary data flows:

- Input stream (ingestion): the connector subscribes to selected MQTT topics exposed by the MASA infrastructure, receiving real-time alert messages and other operational events. These messages represent data updates or contextual signals produced by MASA's monitoring and control systems.

- Output stream (transmission): upon receiving and validating an event, the connector triggers the corresponding API calls to perform lifecycle operations such as Add, Update, or Withdraw a resource in the dataspace. This translation ensures that each MQTT event is reflected as a formal operation within the federated environment.

Through this mechanism, the connector effectively extends the dataspace's perimeter, enabling external IoT-based providers such as MASA (see Section 3.4.1 for details) to publish their data in a standardized and traceable manner. The transmitted events, once registered through the data space, are subsequently notarized on the blockchain by the lifecycle management smart contract, thereby preserving verifiability and integrity across the entire chain of custody.

The MASA connector will go through the onboarding process before being authorized to perform data operations because it is an external entity with regard to the core dataspace. This step ensures that access control procedures are consistently applied and the federated governance model is followed. Once onboarded, the connector will operate autonomously, using its API interface to manage the processes for registered resources. Its modular design effectively positions it as a reusable integration pattern, enabling future scalability toward other external data providers.

From the architectural standpoint, the MASA connector plays a dual role since it is both a technical and governance bridge. Fisrt it connects heterogeneous communication protocols (MQTT and REST APIs) and secondly, it ensures that data, created in operational infrastructures like MASA, enters the dataspace in a compliant and auditable manner.

By completing the interoperability loop between data generation, federation, and notarization, this component thus enhances the previously described blockchain integration and interaction layer. Its development represents the next step in the progressive evolution of the PoC, paving the way for cross-system data flow automation and real-world deployment within the NOUS ecosystem.

4.3 Current Implementation

4.3.1 Sovity Environment

The Sovity sandbox represents the controlled execution environment in which the Proof of Concept operates. It reproduces, in a simplified but fully functional form, the behaviour of a federated dataspace built on the Eclipse Dataspace Connector (EDC) architecture (see a representative snapshot in Figure A.9 in the Appendix). Within this environment, two participants are deployed - Participant 1 and Participant 2 - each representing an autonomous organisation that can act as data provider, data consumer, or both, depending on the exchange scenario.

The services that are deployed for each participant are the following:

- EDC Connector, which authenticates with the Dynamic Attribute Provisioning Service (DAPS) using certificates and OAuth2, publishes resources to the dataspace; then it exposes the data catalogue to other connectors and manages contract negotiation and data transfers.
- EDC Dashboard, a web interface that allows users to interact with the connector; using this tool, users can publish resources, browse catalogues, initiate negotiations, and start a data transfer.
- MinIO Object Storage, acting as the underlying storage layer for data exchange. It provides an S3-compatible API where datasets can be uploaded and later retrieved during the actual transfer.
- Data Exchange Agent, an adapter layer that bridges the EDC Connector and the MinIO storage; it is responsible for the actual data ingestion and delivery during the transfer phase.

The layered structure of the dataspace is reflected in this modular deployment: the data-plane layer (MinIO + Data Exchange Agent) handles the actual physical transfer of content, while connectors oversee governance and control-plane communication.

Functional workflow

Within this sandbox, the PoC follows a complete end-to-end data-exchange workflow, with four main phases.

1. Upload of source data

The data provider first uploads the dataset to its own MinIO instance, specifically to the source-data bucket. This establishes the resource that will later be

referenced by the connector. At this stage the file (e.g., testfile.csv) becomes reachable through an HTTP GET request at a generated MinIO URL. The provider effectively exposes the dataset in a way that the connector can reference, but no dataspace metadata exists yet.

2. Publication of a Data Offer

The provider then registers this dataset in its EDC Connector using the Dashboard interface. On the Create Data Offer page, the user defines:

- Type: REST-API endpoint (reflecting the accessible MinIO object);
- Method: GET;
- URL: the dataset's MinIO URL obtained in the previous step;
- Title: a descriptive, unique name for the asset.

When the user clicks Publish, the connector creates an asset in its internal catalogue, associating the declared data address with optional usage policies. At this point, the resource becomes part of the provider's public catalogue, visible to other participants.

3. Discovery and Contract Negotiation

On the consumer side, the second participant explores available offers through its own Dashboard. By opening the Catalog Browser, the consumer connector queries the provider's DSP endpoint, retrieving and displaying all accessible offers. Once a suitable dataset is identified, the consumer initiates a contract negotiation directly from the Dashboard by selecting the desired offer, requesting negotiation, reviewing terms and confirming the agreement.

When both parties approve, a contract agreement is created and listed under Consuming Contract Agreements on the consumer's interface. This contract governs the subsequent data transfer, enforcing the policies defined by the provider.

4. Data Transfer via Data Exchange Agent

Instead of directly fetching the data from the provider's MinIO, the consumer triggers a transfer mediated by its Data Exchange Agent. From the consumer's Contracts page, the user selects the relevant agreement and initiates the Transfer action, specifying:

- Sink Type: REST-API,
- Method: POST,
- Target URL: the endpoint of the consumer's Data Exchange Agent,

- Headers: appropriate content type (e.g., application/csv).

Once initiated, the transfer request travels through the control plane of the dataspace, is validated against the established contract, and results in the Data Exchange Agent ingesting the dataset. The transferred file is then stored in the consumer's MinIO under the received-data bucket, and the Dashboard shows the transfer as completed in the Transfer History section.

This process successfully demonstrates the dataspace's core promise: data remains within each participant's domain while interoperability, access control, and traceability are achieved through standardised interfaces and protocols.

4.3.2 API-Based Workflow

While the previous section illustrated the end-to-end data exchange workflow as executed through the Sovity Dashboard interfaces, the same sequence of operations can also be performed programmatically by interacting directly with the Management API - administrative interface of the Eclipse Dataspace Connector (EDC) that exposes endpoints of each data space participant. This alternative interaction mode reflects the production-level operation of a dataspace, where participants - rather than manually configuring connectors via the user interface - automate these procedures within their own backend systems or middleware components. In this way, the dataspace becomes machine-operable, allowing complete integration with external services such as the off-chain notarisation and blockchain layers described in the following sections.

The Management API exposes a complete set of REST endpoints that mirror every function available in the Sovity Dashboard. These endpoints allow the registration, configuration, and exchange of resources in a programmatic way, ensuring that every step in the data lifecycle can be executed deterministically and reproducibly. From an architectural perspective, this design decouples the user interface layer (dashboard) from the control plane API, providing flexibility in how each participant interacts with its connector - manually for human-operated testing, or automatically for system-level integration. Using the API to replicate the workflow previously executed in the dashboard provides several advantages:

- Automation and repeatability: workflows can be executed repeatedly without human intervention, allowing end-to-end tests and notarisation pipelines to be run automatically.

- Integration with external components: since API responses are structured and machine-readable, they can be directly consumed by external services - such

as the webhook responsible for canonicalization[1] and hashing.

- Scalability and interoperability: real dataspaces are expected to interoperate through system-to-system communication rather than manual user actions.

The API-based approach therefore aligns the PoC with realistic, production-grade scenarios.

Every step previously described in the sandbox workflow has a direct API equivalent. This parity ensures that the behaviour observed in the dashboard can be faithfully reproduced and automated.

By invoking the corresponding endpoints sequentially, one can reproduce the entire sequence of actions performed through the graphical interface - covering asset publication, policy creation, contract definition and negotiation, and eventual data transfer.

In the Proof of Concept implementation, the API-based interaction serves a dual purpose: (i) it allows the automation of dataspace workflows, transforming what was previously a manual sequence into a reproducible process; (ii) it provides the technical interface for forwarding dataspace events to the external notarisation service.

To achieve this, a curated Postman collection of API requests was employed. The collection acts as a programmable client capable of executing all management operations on the Sovity sandbox, while also embedding custom scripts that extend its behaviour.

Adopting an API-driven workflow extends the scope of the Proof of Concept beyond a user-facing demonstration. It demonstrates that the dataspace connector can serve not only as a human-operated platform but also as an integrable component in larger distributed architectures. In this setup, the Management API acts as the gateway between the dataspace domain and external trust infrastructures. The benefits of this integration can be summarised as follows:

- Technical cohesion: both the dashboard and the API share the same connector logic, guaranteeing that API calls reproduce the exact behaviour of manual operations.
- Auditability and traceability: every request and response exchanged through the API can be logged, timestamped, and later cross-verified with the on-chain records generated by the webhook.

[1]Canonicalization refers to the process of converting a JSON object into a single, deterministic representation by ordering its keys and removing non-semantic variations (such as whitespace or indentation). This ensures that logically equivalent documents always produce the same hash value, enabling reproducible and verifiable notarization on the blockchain.

- Extensibility: the use of open REST interfaces allows new processes - such as incremental validation, policy updates, or transfer audits - to be integrated without altering the connector's core functionalities.

By establishing this API-based automation layer, the PoC bridges the operational domain of the dataspace and the trust domain of the blockchain, preparing the ground for the integrated flow presented in the next section.

4.3.3 End-To-End Flow

This section introduces the workflow representing the complete lifecycle of a notarized dataspace event. It starts from an API call to the dataspace, continues through the off-chain webhook actions, and culminates in the blockchain transaction that permanently records the event.

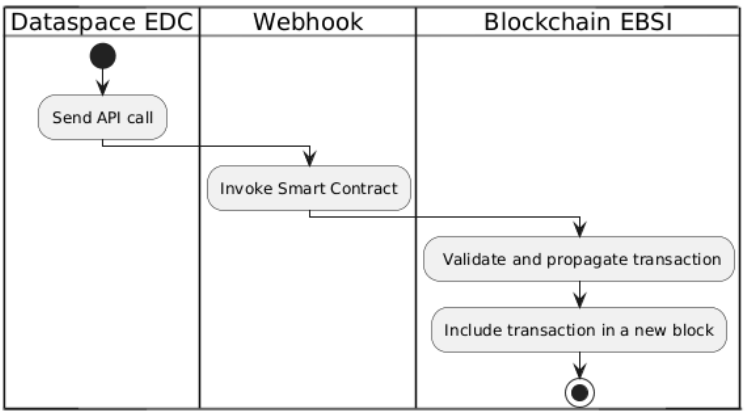

Figure 4.5: Webhook-Based Integration Workflow

The integration between the dataspace and the blockchain constitutes the final stage of the Proof of Concept implementation, where the trust mechanisms of a distributed ledger are coupled with the data governance functions of the dataspace. Figure 4.5 illustrates the end-to-end workflow realized in this phase, showing the interaction among the three core components.

From left to right, the diagram captures the logical progression of each operation:

- API Call - the dataspace participant executes an operation such as Create Asset through the Management API.

- Webhook Invocation - the resulting payload, containing the authoritative metadata of the operation, is automatically sent to the webhook service.

- Blockchain Interaction - the webhook processes the payload, generates deterministic hashes, and invokes a smart contract on the blockchain to record these proofs immutably.

This pipeline transforms every eligible dataspace operation into a verifiable, cryptographically anchored event. It extends the dataspace's internal logging and governance features with external proof of integrity and non-repudiation guarantees.

Figures A.9, A.10, and A.11 provide illustrative views of the dataspace, webhook, and blockchain interfaces implemented within the PoC environment.

API Call

The process begins when a participant performs an operation through the Management API. For example, a data provider might create a new asset in the dataspace. Once the dataspace confirms the successful execution of this operation, the automation logic (implemented through Postman scripts) retrieves the relevant data entities via GET requests to the Management API. The responses from these calls represent the authoritative view of the dataspace state at that moment. They are combined into a single JSON payload and this object constitutes a semantic snapshot of the dataspace event. The payload is then forwarded to the webhook endpoint. This communication marks the boundary between the dataspace control plane and the off-chain notarisation layer.

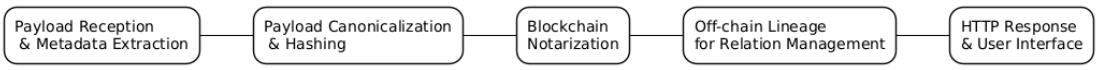

Figure 4.6: Webhook Steps

When the dataspace emits the event payload, the webhook acts as the intermediary layer that translates the dataspace operation into a structured and verifiable record (see Figure 4.6). This component performs a sequence of off-chain actions - focused on processing, normalization, and preparation of metadata - and concludes with an on-chain transaction that permanently anchors the cryptographic proof of the event on the blockchain. Together, these two domains ensure that every dataspace event is both human-readable and cryptographically verifiable.

Event interpretation and metadata extraction

Upon receiving the payload, the webhook first interprets its content to identify the essential descriptive elements of the event. It analyses the incoming JSON

structure to reconstruct a concise summary of the operation: the unique identifier of the asset involved, its version, the author, the title, and any references to other related assets. In this stage, the webhook's purpose is not to transform the data but to summarize from the dataspace message the meaningful information that characterizes the event. The resulting extracted view provides a clear snapshot of what happened in the dataspace and when, establishing the foundation for the notarization step that follows.

Canonicalization and cryptographic hashing

Once the relevant information has been organised, the webhook canonicalizes the entire payload - reordering its structure in a deterministic way - to ensure that identical events always generate the same digital fingerprint. The hash of this canonical representation becomes the unique cryptographic proof of the dataspace event. This process ensures that even if the same event is reproduced in another environment, its resulting hash will match exactly, allowing anyone to verify its integrity independently.

Blockchain notarization

The webhook then interacts with the blockchain and at the current stage of the Proof of Concept, this interaction occurs with a local Ethereum-compatible environment (Ganache). It involves a dedicated smart contract (its internal structure and design rationale are described in detail in the next section) which transmits a minimal but complete set of information: the event type, the identifier of the asset, the computed hash of the payload, and the timestamp of occurrence. The blockchain validates the transaction, includes it in a new block, and produces an immutable transaction hash (txHash) that uniquely represents the notarized event. This identifier - together with the block number, block timestamp, and gas usage - forms the on-chain evidence returned to the webhook and, indirectly, to the dataspace process that originated the event.

Recording of asset relationships and lineage

After the on-chain submission, the webhook enriches the information off-chain by storing references between assets - for example, when one version derives from another or when an asset is a modification of a previous one. This additional layer supports visibility and queryability: it allows lineage graphs to be explored easily, facilitating searches, comparisons, or dependency analysis among dataspace assets. However, the integrity of these relationships does not rely solely on this off-chain record. The relations themselves are already part of the payload that the webhook hashed and notarized during the blockchain step. In other words, even if

the lineage information is exposed in a readable and queryable way for convenience, its complete representation is already contained, through the metadata, within the immutable blockchain proof.

For operational transparency, the webhook provides a web-based interface that visualises events in real time. The Live Logs section traces the main stages of each transaction (sent, mined, gas used), while the Latest Events section presents a structured summary showing the key metadata, blockchain information, and a readable version of the entire payload. This dual view allows the operator to follow both the semantic evolution of the dataspace and the technical evidence of its notarization.

Notarization Smart Contract

Within the proposed Proof of Concept, the DataspaceNotary smart contract constitutes the core blockchain component responsible for ensuring integrity and traceability of operations executed within the dataspace. Its design combines two complementary mechanisms - state storage and event emission - which together establish a dual-layer architecture that supports both on-chain verifiability and off-chain interoperability. Although both mechanisms register the same logical occurrence (a data-related event associated with a specific asset, actor, and timestamp), they serve distinct and complementary purposes.

State storage

Each invocation of the notarize() function creates a new entry within the contract's persistent state by mapping a unique identifier (eventId) to a structured record. This record contains essential metadata such as the event type, the asset identifier, the hash of the associated payload, the timestamp of the operation, the address of the initiating participant, and the corresponding block metadata. Storing these values in the contract's state ensures their permanence and immutability: the information becomes part of the blockchain's canonical ledger and can therefore be queried, verified, and referenced by any participant or other smart contract. This on-chain persistence provides an authoritative and non-repudiable audit trail of activities, enabling the reconstruction of the chronological evolution of an asset and the validation of integrity claims by recomputing payload hashes and verifying the recorded provenance. Because writing to the blockchain state is computationally and economically expensive, the stored information is intentionally limited to metadata that is strictly necessary for evidential and governance purposes. The original data, or payload, remains stored off-chain, thereby preserving confidentiality while allowing the blockchain to serve as a trusted attestation layer.

Event emission

In addition to modifying the state, the smart contract emits an event Notarized that is recorded within the transaction log of the block. Although events are

permanent elements of the blockchain ledger, they do not form part of the contract's storage and cannot be accessed by other smart contracts. Their function is primarily communicative: they constitute an asynchronous and efficient interface through which external systems can be notified of new notarizations in near real time. Off-chain components such as middleware platforms, dashboards, auditing tools, and monitoring systems can subscribe to these event logs to receive updates as soon as they are produced. By indexing specific parameters, events can be efficiently filtered and retrieved by external applications that may, for example, reconstruct the history of all notarizations related to a specific asset or analyse all occurrences of a given operation type. This mechanism substantially enhances observability and allows different actors or services within the dataspace to remain synchronized with the most recent on-chain state without the need for constant polling.

Thus, the architectural decision to combine storage and events strikes a balance between interoperability, cost, and durability. Contracts and auditors can both consult the authoritative, queryable collection of notarisations established by storage. Events support asynchronous, decoupled consumers by offering a scalable distribution channel for state changes. Additionally, the dual path also accomodates different temporal access patterns such as streaming log consumption for operational responsiveness and archival queries against storage for legal reconstruction and compliance. In practice, this design lowers the burden on off-chain components: instead of periodically scanning the entire state, they subscribe to a filtered event stream and, when needed, access individual records in storage to verify their local projections.

A further advantage of the event channel is workflow automation. Because logs are inherently ordered and include block-level timing, off-chain orchestrators can encode reaction rules. These kind of automations could be implemented without expanding on-chain logic, preserving the contract's minimal surface and keeping enforcement where it belongs for dataspace scenarios - within policy engines and connectors that control actual data access. In future implementations, when automation requires stronger guarantees, the on-chain record can include identifiers of the relevant policies or agreements, allowing off-chain enforcement components to reconcile a request with the latest notarized state.

From a system point of view, this pattern also makes solid observability possible. Namely, the events component supports selective indexing (e.g., via log-based indexers or query engines) to produce low-latency, ad hoc views without changing the contract. Storage then provides a reliable reference for reconciliation and dispute resolution. Given an external document, its digest can be recalculated and compared to the on-chain payload hash to confirm that the same content existed at or prior to the notarized time. This is made possible by the strict immutability of recorded hashes, which serves as an anchor for reproducibility. The blockchain provides integrity, time, and provenance, while the contract stores only hashes and

minimal metadata, supporting privacy-by-design. Sensitive contents stay in the object store or connector-controlled endpoints.

Additionally, the design further allows for evolution and specialization without contract proliferation. With an emphasis on properties that are constant across operations (type, asset, content hash, time, actor), the record schema is purposefully brief and orthogonal. When operation-specific information is required, it can be introduced as versioned schemas that are referred to by identifiers kept in the record, or it can be carried off-chain and cryptographically bound through the payload hash. In order to improve indexability and maintain storage efficiency, events can expose extra, fields, specific for certain operations. This approach allows the same contract to notarize multiple lifecycle steps while preserving a single audit trail per asset.

Lastly, assurance and governance are enhanced by this dual mechanism. Internal and external audits, certification procedures, and inter-organizational dispute resolution can all rely on storage-backed records as their evidentiary foundation. Event streams provide operational assurances: after a change, off-chain controls can quickly converge to a consistent state and stakeholders are promptly informed. The end result is a faithful alignment with the dataspace paradigm: the blockchain provides a shared, tamper-evident substrate for integrity, chronology, and attribution, enhanced by an effective and filterable notification surface, but it neither carries the data itself nor tries to centralize enforcement.

The integration of on-chain notarization into the dataspace architecture ultimately results in a set of tangible advantages that extend across technical, organizational, and governance dimensions:

- Data integrity: if an asset is altered, it is detectable through the cryptographic hashing and immutable recording of metadata.
- Provenance and attribution: every record is tied to the blockchain address that initiated the operation, allowing attribution of actions to distinct participants. Beyond identification, the ability to reconstruct a certified and tamper-evident lineage of each asset over time enhances the informational and economic value of data, as its lifecycle becomes demonstrably authentic and verifiable.
- Auditability: event logs and persistent on-chain storage allow for creating a verifiable audit trail that can be examined by internal and external authorities.
- Interoperability and automation: event-driven notifications facilitate to synchronize with off-chain platforms, enabling responsive data governance and automated workflows.
- Scalability and efficiency: the architecture minimizes storage and transaction costs while preserving full traceability, given the storing of only metadata and

hashes on-chain.

- Trust and compliance: notarized entries enhance transparency and accountability across participants by providing technically verifiable evidence of data operations, which can complement existing compliance and certification processes.

4.4 Value Creation and Impact of the Proposed Framework

The previous sections presented the architectural and technical foundations of the proposed blockchain-enabled framework for data lifecycle management within federated data spaces. Having established the logical and operational components of the system, this final section shifts the focus towards the value dimension, thus how the implemented architecture translates into valuable benefits for the involved stakeholders and, more broadly, for the digital ecosystem it supports. The purpose here is to move from a structural perspective to an interpretative one, showing how the technical elements described earlier generate functional and strategic value across multiple levels of the data value chain.

At its core, the proposed framework operationalizes the principles of trust, accountability, and interoperability that underpin the European Data Strategy. By embedding notarization and verifiable data handling within the operational workflows of the data space, the system ensures that each data transaction is traceable, auditable, and verifiably authentic. This shift transforms the data space from a coordination environment based on contractual trust into an infrastructure of technological trust, where reliability is guaranteed through design rather than through institutional assumption.

Beyond ensuring verifiability, this continuous traceability of the data lifecycle constitutes the very source of value creation. The framework demonstrates that a notarized system can effectively record the complete life of a data asset - its creation, usage, transformation, and eventual retirement - under a standardized and immutable structure. Each event associated with the data is certified and time-stamped, allowing future processes to reconstruct its exact trajectory and verify its legitimacy even as data formats, standards, or anonymization requirements evolve. This lifelong certification of data ensures that every derivative or version remains linked to its original, enabling responsible governance and informed action in any subsequent scenario. For instance, if a data asset must later be invalidated or updated, the recorded links allows the identification of all dependent processes, ensuring consistent and transparent propagation of changes throughout the ecosystem. In this way, value emerges from the act of tracking itself - from the

ability to certify not only what a piece of data is, but what it has been and how it has been used.

The integration of the NOUS components within the MASA use case demonstrates this transformation in a concrete, domain-specific scenario. MASA serves as an ideal validation environment due to its inherently multi-actor nature. Through the incorporation of NOUS and its blockchain-integrated framework, the MASA platform evolves from a distributed mobility service network into a trusted, interoperable, and data-driven decision-support environment.

Within this context, value generation manifests along three complementary dimensions: operational efficiency, data integrity and governance assurance, and systemic collaboration enablement. Each dimension corresponds to distinct needs and expectations of the principal stakeholder groups, namely municipal managers, MASA IT managers, and safety or insurance managers, whose activities intersect within the shared mobility ecosystem.

For municipality managers, the framework addresses long-standing challenges in data reliability, interoperability, and evidence-based governance. Municipal authorities are responsible for monitoring traffic flows, ensuring road safety, and identifying high-risk zones in increasingly complex urban networks. Traditionally, these activities rely on fragmented, uncertified, and often incompatible data sources, leading to reactive rather than predictive decision-making. Through the proposed framework, municipalities gain access to certified and notarized data streams that preserve integrity and authenticity throughout their lifecycle. The notarization mechanism ensures that every dataset - whether derived from sensors, cameras, or predictive algorithms as detailed in Section 3.4.1 - can be verified as untampered and traceably linked to its origin. This not only enhances the credibility of urban mobility analytics but also enables regulatory compliance and auditability, both essential for public sector accountability.

Moreover, the integration of NOUS within MASA facilitates AI-assisted predictive safety and real-time monitoring capabilities. The availability of harmonized and trustworthy data across multiple participants allows municipalities to transition from local, siloed data management to systemic, cross-city intelligence. Planners and policymakers can thus make informed decisions grounded in verifiable evidence, improving traffic optimization, risk mitigation, and the design of safety policies for vulnerable road users. In essence, the framework transforms the municipality's role from a passive data consumer into an active orchestrator of interoperable and certified urban mobility intelligence.

For MASA IT Services Managers, value creation occurs primarily through the simplification and automation of data operations. Managing heterogeneous data sources, ensuring interoperability, and maintaining security standards are typically

high-cost tasks. Through automated data harmonization and integrated notarization, the technical workload associated with format conversion, validation, and certification is drastically reduced. NOUS assumes the role of a trusted intermediary that governs access permissions and data certification, while the underlying dataspace infrastructure guarantees interoperability through standardized protocols. This reallocation of responsibilities enables MASA IT teams to focus on higher-value objectives such as service optimization, AI model refinement, and advanced mobility analytics. The result is a more agile operational model, where technical resources are efficiently directed toward enhancing the scientific and societal value of data.

For safety and insurance managers, the impact is particularly evident in the domain of risk evaluation and incident management. These actors depend on reliable, privacy-compliant data to assess liabilities, manage claims, and design insurance products aligned with urban safety goals. Prior to the implementation of the framework, access to consistent and verifiable safety data was constrained by a lack of certification across municipalities and service providers. With the introduction of blockchain notarization, every relevant data point - whether it concerns accident reports, vehicle trajectories, or environmental conditions - can be cryptographically validated and time-stamped. This creates a foundation for trustworthy, privacy-preserving, and legally defensible safety analytics. Additionally, the interoperability layer of the dataspace allows safety managers to seamlessly access and share data across different organizational boundaries under explicit usage policies. This feature significantly reduces the latency typically associated with liability attribution and claim resolution. The result is a data ecosystem where safety information becomes a shared and monetizable asset, available for collaborative innovation among municipalities, insurers, and technology providers.

The collective impact of these improvements extends beyond individual stakeholders. At the ecosystem level, the integration of NOUS and MASA demonstrates how federated data infrastructures can evolve into self-sustaining trust ecosystems, where each participant contributes to and benefits from a continuous cycle of certified data exchange. The blockchain layer reinforces this ecosystemic value by embedding verifiability and permanence into its operations, transforming digital transactions into immutable records of accountability. Such traceable operations foster long-term reliability and enable transparent auditing mechanisms that are particularly relevant for public-sector governance and cross-domain compliance. Furthermore, the implemented framework establishes a replicable blueprint for interoperability and trust management across different European data spaces. The MASA use case, while domain-specific, embodies a generalizable approach to the integration of blockchain and data space technologies. Its layered architecture can be adapted to other contexts where secure, accountable, and sovereign data sharing

is required.

By enabling verifiable and standardized data exchange, the framework contributes to the broader European objective of establishing a trusted digital single market for data. It operationalizes regulatory frameworks such as DGA and aligns with ongoing efforts to ensure that data sovereignty is preserved while promoting innovation and cross-border collaboration. In this sense, the MASA-NOUS integration can be seen not only as a technical proof of concept but also as a policy-aligned demonstration of trusted data federation, providing empirical evidence of how decentralized technologies can serve public-interest goals.

Ultimately, the value creation and impact of the proposed framework lie in its capacity to translate complex technological constructs into tangible societal and operational benefits. It demonstrates that trust can be engineered into data ecosystems, making reliability an intrinsic feature and allowing the ecosystem to move toward a new paradigm of verifiable, sovereign, and sustainable data collaboration.

Chapter 5

Conclusions and Future Work

This thesis has investigated the integration of blockchain technology within federated data space architectures, developing both a conceptual framework and a working Proof of Concept (PoC) that demonstrate how notarization, traceability, and accountability can be technologically enforced across the data lifecycle. Through the methodological convergence of the NOUS architectural framework and the MASA use case, the thesis has validated the feasibility and value of embedding blockchain-enabled trust mechanisms into European data ecosystems, establishing a concrete and solid foundation for infrastructures based on verifiable trust and data traceability.

From a conceptual point of view, the thesis has articulated how blockchain can complement existing governance models in data spaces by providing immutable and verifiable records of data operations. Data spaces, by design, rely on federated governance where participants interact under shared rules, maintaining autonomy and sovereignty over their data. Blockchain introduces a technological enforcement of accountability by notarizing every relevant operation on an immutable ledger. This integration bridges the gap between normative compliance and technical verifiability, aligning with the European Data Strategy's emphasis on transparency, sovereignty, and interoperability.

The technical implementation validates several critical design decisions. The smart contract architecture, combining state storage for authoritative on-chain records with event emission for asynchronous notification, achieves an effective balance between verifiability, cost efficiency, and interoperability. By storing only cryptographic hashes and minimal metadata on-chain, the system respects privacy-by-design principles and complies with data minimization requirements required by European data protection regulations. This dual mechanism enables

both real-time operational responsiveness through event streams and retrospective auditability through persistent state queries, supporting diverse temporal access patterns required by different stakeholders.

Equally important, the thesis demonstrated that notarization does not simply serve compliance or auditability purposes, it actively generates value. By enabling verifiable lifecycle tracking, the framework turns data into a certifiable and durable asset. Each dataset carries with it a traceable history of transformations and validations, enhancing its informational reliability and economic value. This aligns with the principle that the true value of data emerges from its reuse under verified provenance.

The synergistic integration with the MASA provides concrete evidence of this principle, showing the framework's applicability in real-world scenarios characterized by multiple autonomous actors, heterogeneous data sources, and stringent requirements for both performance and trust. The integration demonstrates how a blockchain-enabled data space can transform urban mobility data management from fragmented, uncertified information flows into a coordinated ecosystem where data integrity, provenance, and authenticity are cryptographically guaranteed throughout the entire lifecycle. The operational benefits observed across different stakeholder groups illustrate how technical trust mechanisms translate into tangible organizational and societal value.

Beyond the immediate technical achievements, this work contributes to the broader discussion on trust architectures for federated digital ecosystems. Blockchain integration need not introduce complexity at the user level or require participants to manage cryptographic credentials directly. Instead, blockchain functions as an infrastructural trust layer, operating transparently beneath governance and operational interfaces while providing independently verifiable proofs when needed for auditing, compliance verification, or dispute resolution. This architectural positioning resolves a fundamental tension in federated systems: the need to establish trust without centralizing control or imposing heavy operational burdens on participants.

The framework also advances understanding of how technical trust and institutional trust can complement rather than compete with one another. Data space governance remains the primary mechanism for defining participation rules, access policies, and usage conditions, while blockchain provides cryptographic enforcement and verification of those policies' execution. This division of responsibilities enables each layer to fulfill its function optimally, with governance frameworks remaining flexible and adaptable while the blockchain substrate ensures consistent and verifiable execution of agreed-upon rules.

The current implementation, while functional and demonstrative of core concepts, represents an early-stage prototype that establishes the foundation for substantial future development. The roadmap for advancing this work encompasses three

primary evolutionary axes that will progressively enhance the system's maturity, scalability, and production readiness.

The blockchain incremental validation framework defines the evolutionary path from the current TRL 4 implementation through Hyperledger Besu deployment at TRL 5 - enabling validation under realistic consensus conditions and operational procedure establishment - toward full EBSI integration at TRL 6, which will address institutional governance compliance, Decentralized Identifier frameworks, and cross-border interoperability validation. The evolution of the interaction layer addresses fundamental scalability and resilience limitations of the current webhook-based integration. The planned transition to message-oriented middleware enables asynchronous processing, delivery guarantees, message persistence for audit reconstruction, and horizontal scalability independent of dataspace connector capacity. The data provider integration layer completes end-to-end automation by eliminating manual intervention points. The planned MASA connector will bridge MQTT-based MASA infrastructure and standardized dataspace interfaces, enabling real-time propagation of mobility events from sensor networks through federated data spaces to immutable blockchain records.

Beyond these technical evolution axes, future work should address the development of a data economy assessment framework to evaluate value creation mechanisms, transaction cost structures, and sustainable governance models for cost allocation and incentive alignment across blockchain-enabled data space participants.

In conclusion, this thesis establishes both the conceptual foundation and practical demonstration of blockchain-enhanced data lifecycle management in federated data spaces. The implemented framework shows that distributed ledger technology can meaningfully enhance trust and accountability in data sharing ecosystems while respecting sovereignty principles and regulatory requirements central to the European digital strategy. The progressive development roadmap provides a clear path toward production-grade deployment. As European data spaces mature toward operational reality, the mechanisms developed and validated in this work offer a viable approach to technically enforceable trust that complements institutional governance, supporting the realization of sovereign, accountable, and verifiable digital infrastructures.

Appendix A

Appendix

A.1 GitHub Link

The source code related to the Proof of Concept (PoC) developed for this thesis is available at: GitHub Project Repository.

A.2 Glossary

API Application Programming Interface enabling interaction between software systems over a network.

Asset Resource (e.g., dataset) registered in a provider's catalogue and governed by policies and contracts.

Block Unit of transactions appended to a blockchain, with metadata such as timestamp, number, and receipts.

Blockchain Distributed ledger providing immutability and shared verifiability for recorded events.

Canonicalization Deterministic serialization of data (e.g., stable JSON key ordering) so identical inputs yield the same hash.

Catalogue Published list of data offers discoverable by other participants.

Connector Middleware component linking an organization to the dataspace for discovery, policy enforcement, negotiation, and transfer.

Control Plane Coordination layer managing catalogues, negotiations, and policy checks (separate from the data path).

Contract Agreement Binding outcome of negotiation defining conditions and policies for a consumer's access to an asset.

Cryptographic Hash Fixed-length digest used to detect tampering and to anchor notarization on-chain.

Data Offer Description of how an asset can be accessed (type, method, endpoint) and under which usage policies.

Data Plane Layer performing the actual data transfer between participants.

Dataspace Federated ecosystem where autonomous participants share data under shared governance and interoperability principles.

DAPS Dynamic Attribute Provisioning Service issuing short-lived tokens with dynamic attributes for connector authentication.

DGA Data Governance Act, EU regulation establishing trustworthy frameworks for data sharing and reuse.

EBSI European Blockchain Services Infrastructure for trusted cross-border digital services.

EDC Eclipse Dataspace Connector enabling federated data exchange with policy enforcement and contract negotiation.

Ganache Local Ethereum-compatible blockchain used for testing and validating transactions.

GDPR General Data Protection Regulation governing personal data processing and protection in the EU.

Governance Authority Entity managing onboarding, credentialing, and policy rules for participants in data spaces.

IDS International Data Spaces, reference architecture for secure and sovereign data sharing.

MASA Modena Automotive Smart Area, smart-city environment providing mobility datasets for the PoC.

MinIO S3-compatible object storage used to host datasets involved in transfers.

MQTT Lightweight publish/subscribe messaging protocol commonly used for IoT communication.

Notarization Recording hashed payload metadata on a blockchain to provide immutable, verifiable evidence of an event.

Payload Structured data body transmitted in a request or event, containing the information to be processed, hashed, or notarized.

PoC Proof of Concept demonstrating feasibility of the proposed blockchain-dataspace workflow.

Policy Usage rule bound to an asset and enforced by the connector during negotiation and access.

Smart Contract Program deployed on a blockchain that maintains state and emits events to notarize operations.

Sovity Sandbox environment based on the Eclipse Dataspace Connector used to instantiate the federated PoC.

Transaction Hash Unique identifier of a blockchain transaction confirming inclusion in a block.

TRL Technology Readiness Level indicating system maturity from prototype to deployment.

UML Unified Modeling Language used to diagram processes and system interactions.

Webhook HTTP endpoint receiving lifecycle events, performing hashing/canonicalization, and invoking the smart contract for notarization.

A.3 Diagrams

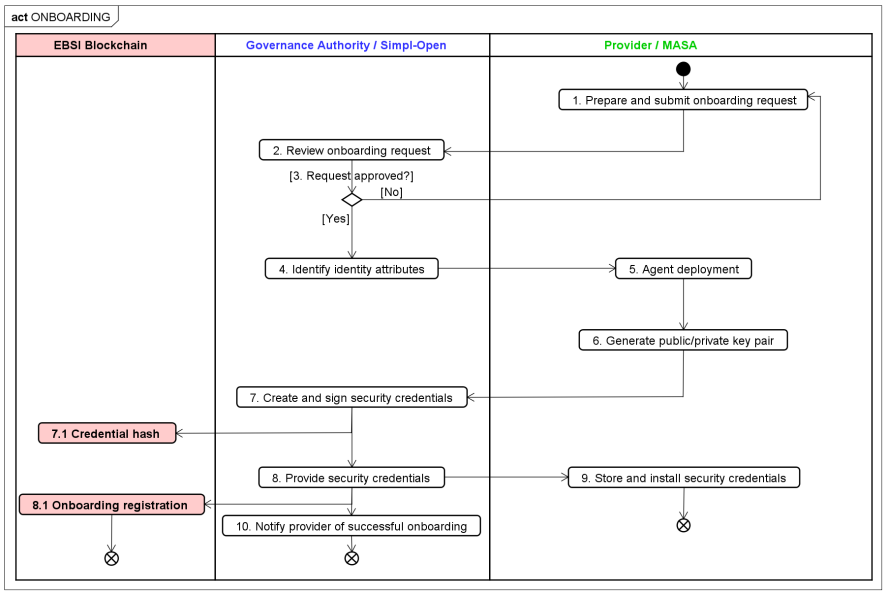

Figure A.1: UML Diagram: Onboarding Operation

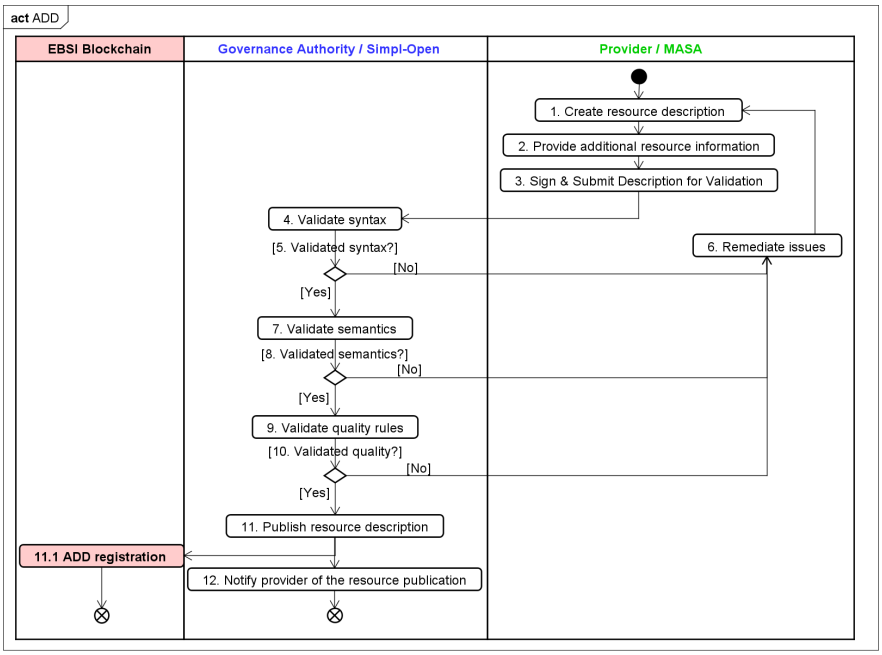

Figure A.2: UML Diagram: Add Operation

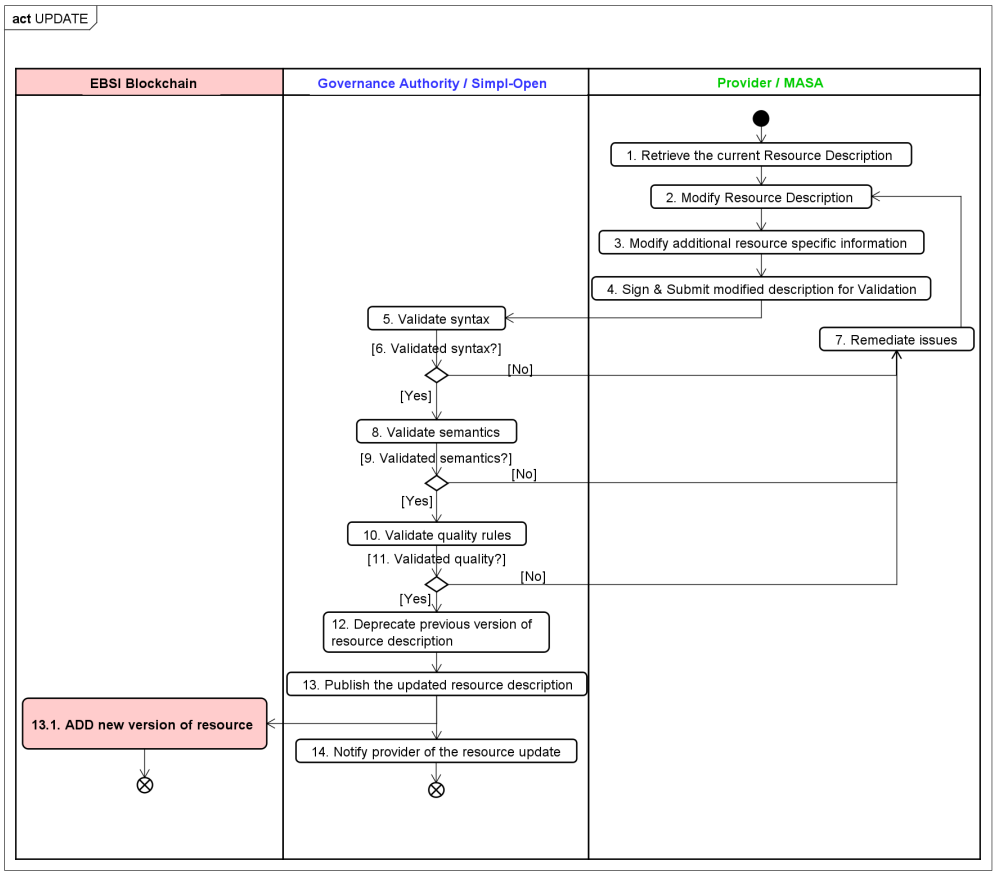

Figure A.3: UML Diagram: Update Operation

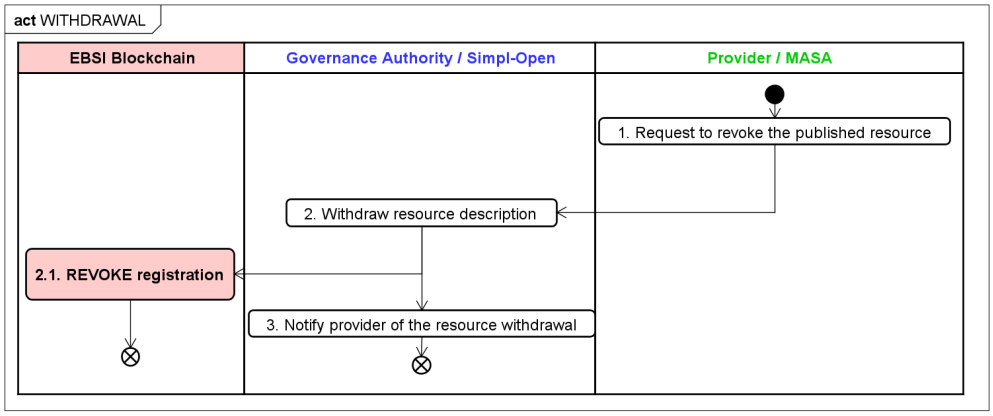

Figure A.4: UML Diagram: Withdrawal Operation

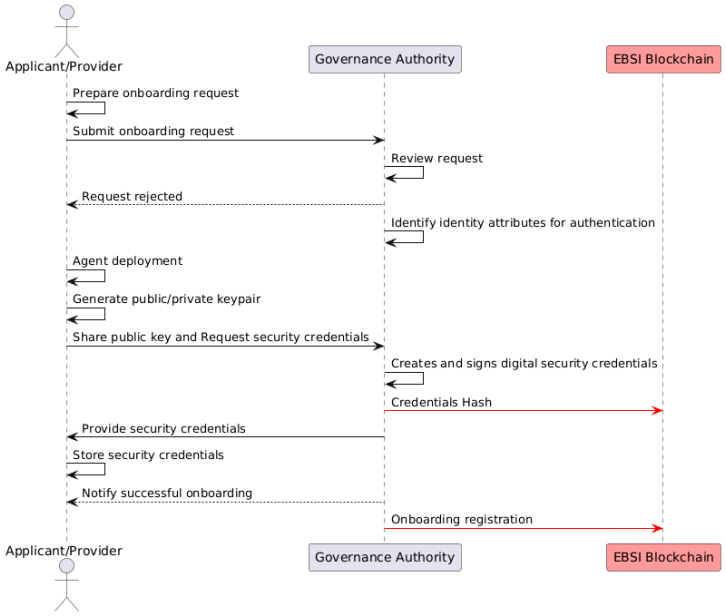

Figure A.5: Sequence Diagram: Onboarding Operation

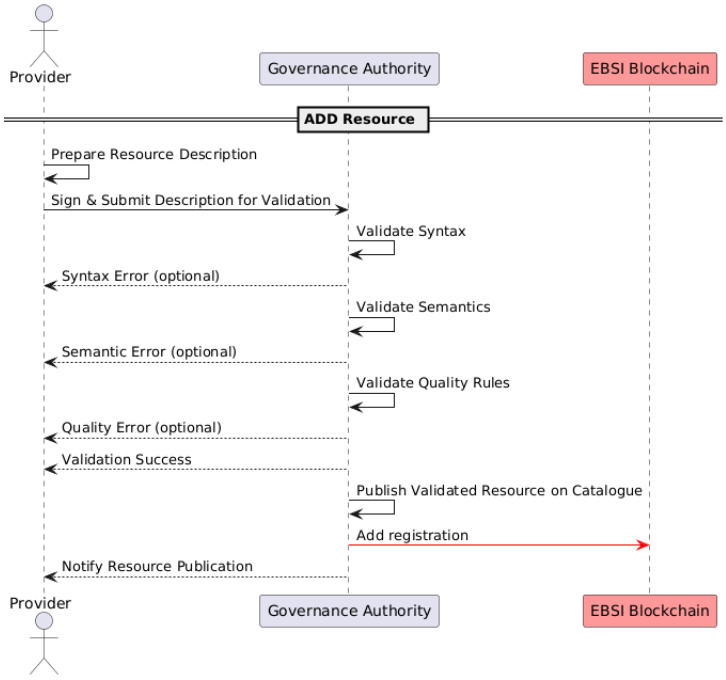

Figure A.6: Sequence Diagram: Add Operation

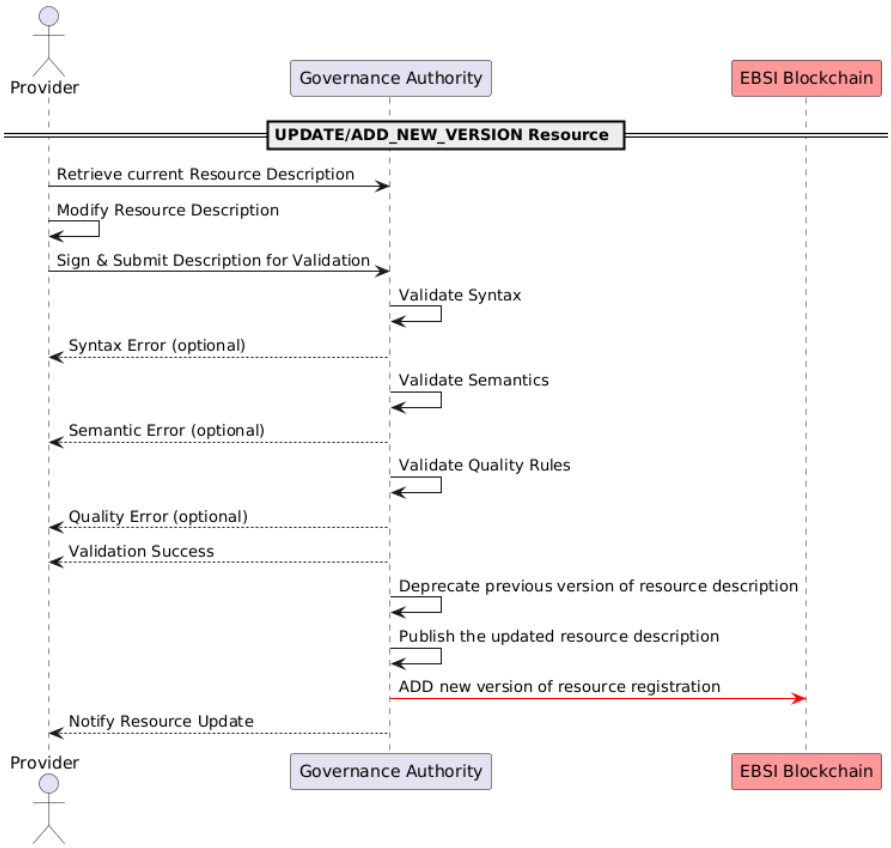

Figure A.7: Sequence Diagram: Update Operation

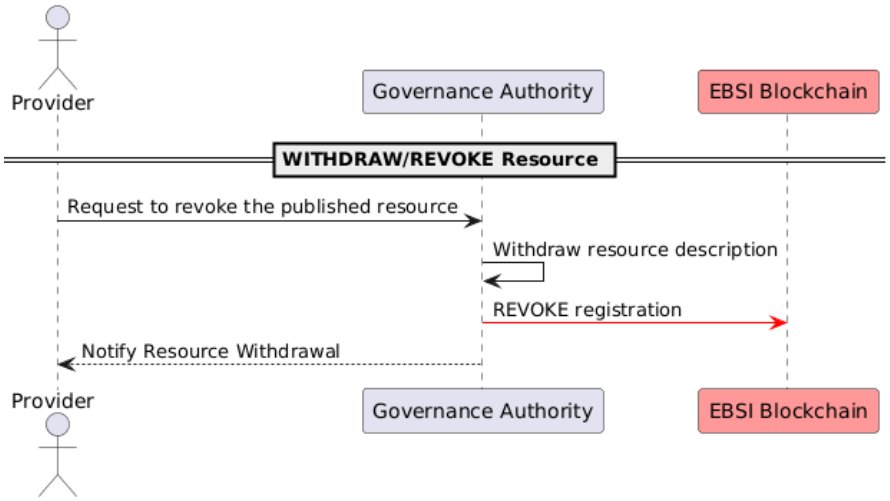

Figure A.8: Sequence Diagram: Withdrawal Operation

A.4 PoC Current Implementation

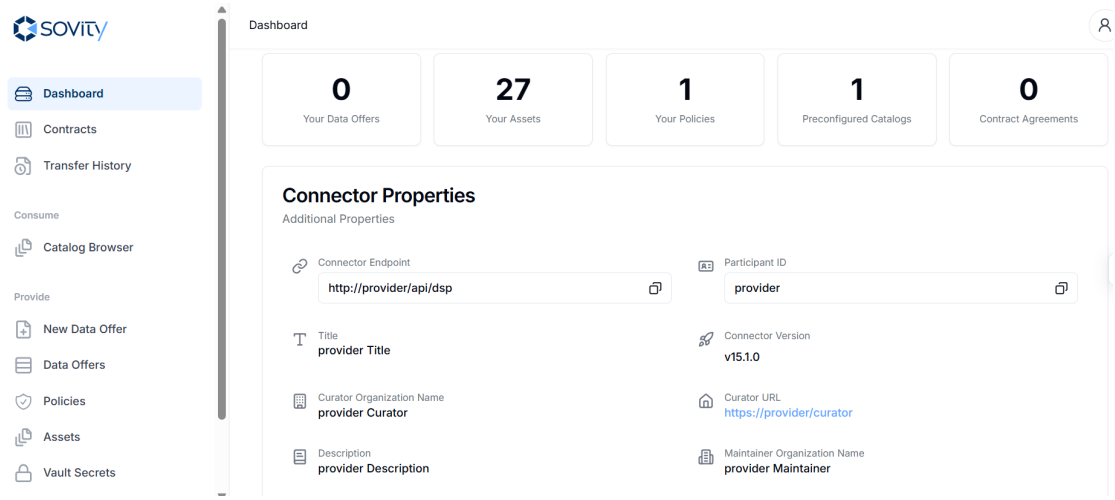

Figure A.9: Sovity Sandbox Dashboard

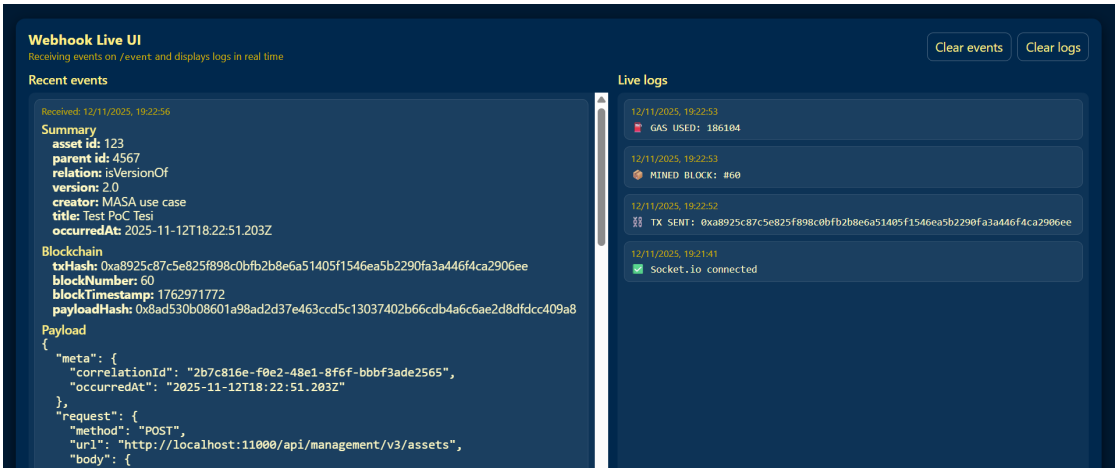

Figure A.10: Webhook

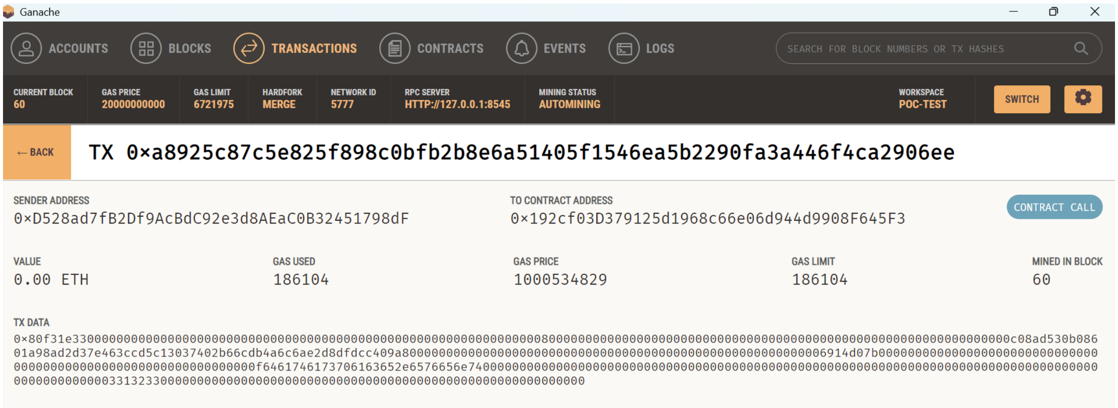

Figure A.11: Ganache Transaction

Bibliography

[1] Ilka Jussen, Julia Schweihoff, Valentin Dahms, Frederik Möller, and Boris Otto. «Data Sharing Fundamentals: Definition and Characteristics». In: *Proceedings of the 56th Hawaii International Conference on System Sciences*. HICSS. Hawaii International Conference on System Sciences, 2023. DOI: 10.24251/hicss.2023.452. URL: http://dx.doi.org/10.24251/HICSS.2023.452 (cit. on pp. 5, 6).

[2] Boris Otto. «The Evolution of Data Spaces». In: *Designing Data Spaces*. Springer International Publishing, 2022, pp. 3–15. ISBN: 9783030939755. DOI: 10.1007/978-3-030-93975-5_1. URL: http://dx.doi.org/10.1007/978-3-030-93975-5_1 (cit. on pp. 5, 6, 9).

[3] Michael Franklin, Alon Halevy, and David Maier. «From databases to dataspaces: a new abstraction for information management». In: *SIGMOD Rec.* 34.4 (Dec. 2005), pp. 27–33. ISSN: 0163-5808. DOI: 10.1145/1107499.1107502. URL: https://doi.org/10.1145/1107499.1107502 (cit. on p. 6).

[4] Alon Halevy, Michael Franklin, and David Maier. «Principles of dataspace systems». In: *Proceedings of the Twenty-Fifth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*. PODS '06. Chicago, IL, USA: Association for Computing Machinery, 2006, pp. 1–9. ISBN: 1595933182. DOI: 10.1145/1142351.1142352. URL: https://doi.org/10.1145/1142351.1142352 (cit. on pp. 6, 8, 10).

[5] European Commission. *European Data Strategy*. URL: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en (cit. on pp. 6, 39, 40).

[6] Manlio Bacco, Alexander Kocian, Stefano Chessa, Antonino Crivello, and Paolo Barsocchi. «What are data spaces? Systematic survey and future outlook». In: *Data in Brief* 57 (Dec. 2024), p. 110969. ISSN: 2352-3409. DOI: 10.1016/j.dib.2024.110969. URL: http://dx.doi.org/10.1016/j.dib.2024.110969 (cit. on pp. 6–9, 12–14).

[7] Johannes Theissen-Lipp, Max Kocher, Christoph Lange, Stefan Decker, Alexander Paulus, André Pomp, and Edward Curry. «Semantics in Dataspaces: Origin and Future Directions». In: *Companion Proceedings of the ACM Web Conference 2023*. WWW 23. ACM, Apr. 2023, pp. 1504–1507. DOI: 10.1145/3543873.3587689. URL: http://dx.doi.org/10.1145/3543873.3587689 (cit. on pp. 9–12).

[8] *The Data Space Manifesto*. International Data Spaces Association, 2025 (cit. on pp. 9–11).

[9] GO FAIR Foundation. *FAIR Principles*. 2025. URL: https://www.go-fair.org/fair-principles/ (cit. on p. 11).

[10] International Data Spaces Association. *International Data Spaces Association - Home*. URL: https://internationaldataspaces.org/ (cit. on pp. 12, 17).

[11] International Data Spaces Association. *The IDS Reference Architecture Model*. URL: https://internationaldataspaces.org/offers/reference-architecture/ (cit. on pp. 12, 17).

[12] Heinrich Pettenpohl, Markus Spiekermann, and Jan Ruben Both. «International Data Spaces in a Nutshell». In: *Designing Data Spaces*. Springer International Publishing, 2022, pp. 29–40. ISBN: 9783030939755. DOI: 10.1007/978-3-030-93975-5_3. URL: http://dx.doi.org/10.1007/978-3-030-93975-5_3 (cit. on pp. 13–16).

[13] Alberto Montero Fernández et al. «A catalyst for European cloud services in the era of data spaces, high-performance and edge computing: NOUS». In: *Proceedings of the 4th Eclipse Security, AI, Architecture and Modelling Conference on Data Space*. eSAAM 2024. ACM, Oct. 2024, pp. 1–9. DOI: 10.1145/3685651.3686660. URL: http://dx.doi.org/10.1145/3685651.3686660 (cit. on pp. 17, 40–43, 45).

[14] NOUS Project. *NOUS - Empowering Europe's Data Future 2025*. 2025. URL: https://nous-project.eu/ (cit. on pp. 17, 40–42).

[15] FIWARE. *FIWARE, the Open Source Platform for Our Smart Digital Future*. 2025. URL: https://www.fiware.org/ (cit. on p. 18).

[16] Data Spaces Support Centre. *Designing and delivering the European single market for data*. 2025. URL: https://dssc.eu/ (cit. on p. 18).

[17] Gaia-X. *Gaia-X Framework*. 2025. URL: https://gaia-x.eu/gaia-x-framework/ (cit. on p. 18).

[18] Gaia-X European Association for Data and Cloud AISBL. *Lighthouse Projects - Gaia-X: a federated secure data infrastructure*. 2025. URL: https://gaia-x.eu/community/lighthouse-projects/ (cit. on p. 18).

[19] EOSC Association. *EOSC Association - Advancing Open Science in Europe*. 2025. URL: https://eosc.eu/ (cit. on p. 19).

[20] Catena-X. *Catena-X - Your automotive network*. 2025. URL: https://catena-x.net/ (cit. on p. 19).

[21] Mobility Data Space. *Mobility Data Space: the data space for future mobility*. 2025. URL: https://mobility-dataspace.eu/ (cit. on p. 19).

[22] SCSN. *Smart Connected Supplier Network*. 2025. URL: hhttps://smart-connected.nl/en/about-scsn/how-it-works (cit. on p. 19).

[23] EONA-X. *Data sharing for transport, mobility and tourism*. 2025. URL: https://eona-x.eu/ (cit. on p. 20).

[24] I. Bashir. *Mastering Blockchain*. Packt Publishing, 2017. ISBN: 9781787125445. URL: https://books.google.it/books?id=dMJbMQAACAAJ (cit. on pp. 20, 23, 24, 26–28).

[25] Changhao Zhu, Junzhe Li, Ziyue Zhong, Cong Yue, and Meihui Zhang. «A Survey on the Integration of Blockchains and Databases». In: *Data Science and Engineering* 8.2 (Apr. 2023), pp. 196–219. ISSN: 2364-1541. DOI: 10.1007/s41019-023-00212-z. URL: http://dx.doi.org/10.1007/s41019-023-00212-z (cit. on p. 20).

[26] Dylan Yaga, Peter Mell, Nik Roby, and Karen Scarfone. *Blockchain technology overview*. Oct. 2018. DOI: 10.6028/nist.ir.8202. URL: http://dx.doi.org/10.6028/NIST.IR.8202 (cit. on pp. 20–27, 29, 30).

[27] W. Diffie and M. Hellman. «New directions in cryptography». In: *IEEE Transactions on Information Theory* 22 (1976). ISSN: 1557-9654. DOI: 10.1109/tit.1976.1055638. URL: http://dx.doi.org/10.1109/TIT.1976.1055638 (cit. on p. 20).

[28] David Chaum. «Blind Signatures for Untraceable Payments». In: *Advances in Cryptology*. Springer US, 1983, pp. 199–203. ISBN: 9781475706024. DOI: 10.1007/978-1-4757-0602-4_18. URL: http://dx.doi.org/10.1007/978-1-4757-0602-4_18 (cit. on p. 21).

[29] Stuart Haber and W. Scott Stornetta. «How to time-stamp a digital document». In: *Journal of Cryptology* 3.2 (Jan. 1991), pp. 99–111. ISSN: 1432-1378. DOI: 10.1007/bf00196791. URL: http://dx.doi.org/10.1007/BF00196791 (cit. on p. 21).

[30] Dave Bayer, Stuart Haber, and W. Scott Stornetta. «Improving the Efficiency and Reliability of Digital Time-Stamping». In: *Sequences II*. Ed. by Renato Capocelli, Alfredo De Santis, and Ugo Vaccaro. New York, NY: Springer New York, 1993, pp. 329–334. ISBN: 978-1-4613-9323-8 (cit. on p. 21).

[31] Satoshi Nakamoto. «Bitcoin: A Peer-to-Peer Electronic Cash System». In: *SSRN Electronic Journal* (May 2008). URL: http://www.bitcoin.org/bitcoin.pdf (cit. on pp. 21, 24, 27).

[32] Alberto Amico et al. «BOTQUAS: Blockchain-based Solutions for Trustworthy Data Sharing in Sustainable and Circular Economy». In: *2024 50th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. IEEE, Aug. 2024, pp. 507–510. DOI: 10.1109/seaa64295.2024.00083. URL: http://dx.doi.org/10.1109/SEAA64295.2024.00083 (cit. on p. 22).

[33] Nick Szabo. «Formalizing and Securing Relationships on Public Networks». In: *First Monday* 2.9 (Sept. 1997). DOI: 10.5210/fm.v2i9.548. URL: https://firstmonday.org/ojs/index.php/fm/article/view/548 (cit. on p. 29).

[34] Vittorio Capocasale and Guido Perboli. «Standardizing Smart Contracts». In: *IEEE Access* 10 (2022), pp. 91203–91212. DOI: 10.1109/ACCESS.2022.3202550 (cit. on pp. 29–31).

[35] Maria Elena Bruni, Vittorio Capocasale, Marco Costantino, Stefano Musso, and Guido Perboli. «Decentralizing Electric Vehicle Supply Chains: Value Proposition and System Design». In: *2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC)*. 2023, pp. 1756–1761. DOI: 10.1109/COMPSAC57700.2023.00271 (cit. on p. 30).

[36] Tharaka Mawanane Hewa, Yining Hu, Madhusanka Liyanage, Salil S. Kanhare, and Mika Ylianttila. «Survey on Blockchain-Based Smart Contracts: Technical Aspects and Future Research». In: *IEEE Access* 9 (2021), pp. 87643–87662. ISSN: 2169-3536. DOI: 10.1109/access.2021.3068178. URL: http://dx.doi.org/10.1109/ACCESS.2021.3068178 (cit. on p. 30).

[37] Vittorio Capocasale, Danilo Gotta, and Guido Perboli. «Comparative analysis of permissioned blockchain frameworks for industrial applications». In: *Blockchain: Research and Applications* 4.1 (2023), p. 100113. ISSN: 2096-7209. DOI: https://doi.org/10.1016/j.bcra.2022.100113. URL: https://www.sciencedirect.com/science/article/pii/S2096720922000549 (cit. on p. 31).

[38] Guido Perboli, Stefano Musso, and Mariangela Rosano. «Blockchain in Logistics and Supply Chain: A Lean Approach for Designing Real-World Use Cases». In: *IEEE Access* 6 (2018), pp. 62018–62028. DOI: 10.1109/ACCESS.2018.2875782 (cit. on p. 31).

[39] Mohamed Elassy, Mohammed Al-Hattab, Maen Takruri, and Sufian Badawi. «Intelligent transportation systems for sustainable smart cities». In: *Transportation Engineering* 16 (2024). ISSN: 2666-691X. DOI: 10.1016/j.treng.2024.100252. URL: http://dx.doi.org/10.1016/j.treng.2024.100252 (cit. on pp. 32–37).

[40] Murali Krishna Pasupuleti. «Smart Mobility: Transforming Roads with Advanced V2X Communication and Connected Vehicle Networks». In: *Connected Mobility: Advancing V2X Communication for Safer and Smarter Roads*. National Education Services, Dec. 2024, pp. 56–72. ISBN: 9788198246004. DOI: 10.62311/nesx/46004. URL: http://dx.doi.org/10.62311/nesx/46004 (cit. on pp. 32–37).

[41] Haiping Si, Weixia Li, Qingyi Wang, Haohao Cao, Fernando Bacao, and Changxia Sun. «A secure cross-domain interaction scheme for blockchain-based intelligent transportation systems». In: *PeerJ Computer Science* 9 (Nov. 2023). DOI: 10.7717/peerj-cs.1678. URL: https://peerj.com/articles/cs-1678/ (cit. on pp. 32, 34–37).

[42] European Commission. *Cloud Computing*. 2025. URL: https://digital-strategy.ec.europa.eu/en/policies/cloud-computing (cit. on pp. 38, 39).

[43] European Union. *EU Cloud and AI Act*. 2025. URL: https://www.eu-cloud-ai-act.com/ (cit. on p. 39).

[44] Enrico Rossini, Marcello Pietri, Roberto Cavicchioli, Marco Picone, Marco Mamei, Roberto Querio, Laura Colazzo, and Roberto Procopio. «5G MEC Architecture for Vulnerable Road Users Management Through Smart City Data Fusion». In: *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*. ACM MobiCom '23. Madrid, Spain: Association for Computing Machinery, 2023. ISBN: 9781450399906. DOI: 10.1145/3570361.3614070. URL: https://doi.org/10.1145/3570361.3614070 (cit. on pp. 45, 47).

[45] Modena Automotive Smart Area. *MASA - A living lab for automated driving*. 2025. URL: https://www.automotivesmartarea.it/ (cit. on pp. 46, 51).

[46] Enrico Rossini, Marcello Pietri, Marco Picone, Carlo Augusto Grazia, and Marco Mamei. «Vulnerable Road Users Accident Prevention via Smart City Data Fusion: Experimental Evaluation of a 5G MEC Architecture». In: *2024 22nd International Symposium on Network Computing and Applications (NCA)*. 2024, pp. 37–44. DOI: 10.1109/NCA61908.2024.00018 (cit. on pp. 46–50).

[47] European Commission. *European Blockchain Services Infrastructure (EBSI)*. URL: https://simpl-programme.ec.europa.eu/book-page/simpl-mission-and-vision (cit. on pp. 54, 56).