Data & Knowledge Engineering 162 (2026) 102539

Contents lists available at ScienceDirect

Data & Knowledge Engineering

journal homepage: www.elsevier.com/locate/datak

Exploring cutting-edge data ecosystems: A comprehensive analysis

Ioannis Chrysakis [a,b,c,*], David Chaves-Fraga [d,b], Giorgos Flouris [e], Erik Mannens [a], Anastasia Dimou [b,f,g]

[a] IDLab, Department of Electronics and Information Systems, Ugent, imec, Technologiepark-Zwijnaarde 126, B-9052 Gent, Belgium
[b] DTAI, Department of Computer Science, KU Leuven, Celestijnenlaan 200A, box 2402, 3001 Leuven, Belgium
[c] Netcompany S.A., Research and Innovation Development Department, 2b Rue Nicolas Bové, L-1253, Luxembourg
[d] CiTIUS, Universidade de Santiago de Compostela, Rúa de Jenaro de la Fuente, s/n, 15705 Santiago de Compostela, A Coruña, Spain
[e] FORTH, Institute of Computer Science, N. Plastira 100, Vasilika Vouton, 70013, Heraklion, Greece
[f] KU Leuven – Leuven.AI, Celestijnenlaan 200A, 3001 Leuven, Belgium
[g] Flanders Make, Gaston Geenslaan 8, 3001, Leuven), Belgium

ARTICLE INFO

Keywords:
Data ecosystems
Data spaces
Data sharing
Data economy

ABSTRACT

Data-driven innovation has recently changed the mindset in data sharing from centralized architectures and monolithic data exploitation by data providers (data platforms) to decentralized architectures and different data sharing options among all involved participants (data ecosystems). Data sharing is further strengthened through the establishment of several legal frameworks (e.g., European Strategy for Data, Data Act, Data Governance Act) and the emerging initiatives that provide the means to build data ecosystems, which is evident in the formulated communities, established use cases, and the technical solutions. However, the data ecosystems have not been thoroughly studied so far. The differences between the various data ecosystems are not clear, making it hard to choose the most suitable for each use case, negatively impacting their adoption. Since the domain is growing fast, a review of the state-of-the-art data ecosystem initiatives is needed to analyze what each initiative offers, identify collaboration prospects, and highlight features for improvement and open research topics. In this paper, we review the state-of-the-art data ecosystem initiatives, describe their innovative aspects, compare their technical and business features, and identify open research challenges. We aim to assist practitioners in choosing the most suitable data ecosystem for their use cases and scientists to explore emerging research opportunities. Furthermore, we will provide a framework that outlines the key criteria for evaluating these initiatives, ensuring that stakeholders can make informed decisions based on their specific needs and objectives. By synthesizing our findings, we hope to foster a deeper understanding of the evolving landscape of data ecosystems and encourage further advancements in this critical field.

1. Introduction

The European Commission recently announced several legal frameworks to facilitate data sharing across different sectors. For instance, the Data Governance Act [1] and the European Strategy for Data[2] boost the development of data-sharing systems. When these data-sharing systems apply decentralized architectures, the involved participants benefit from data reuse and customized data value propositions [3,4]. A *data ecosystem* is defined as a complex socio-technical network that enables collaboration between

* Corresponding author.
E-mail address: ioannis.chrysakis@ugent.be (I. Chrysakis).

https://doi.org/10.1016/j.datak.2025.102539
Received 9 January 2024; Received in revised form 8 April 2025; Accepted 24 November 2025
Available online 2 December 2025

different autonomous actors to share and exploit data [5–7]. Recently, several initiatives appeared with methodologies for building data ecosystems, providing at least an architecture and technical solutions to enable collaborative data sharing [8,9].

The first step to achieving a data ecosystem is to establish a clear framework to support data sharing, the so-called Data Spaces [10]. A data space does not require physical data integration or a common schema [11]. To the contrary, it refers to a virtual or conceptual environment where data is stored, managed, and made available for various purposes [12]. EU promotes common European Data Spaces, e.g., in European Strategy for Data [2], ensuring that more data will become available for use in the economy and society while keeping the companies and individuals who generate the data in control [12,13]. Therefore, we notice a substantial growth of data spaces across diverse sectors (e.g., Mobility [14–16], Energy [17], etc.).

Since the domain of data ecosystems grows so fast, a comparative analysis of the established initiatives is needed, as each initiative covers different needs and fits different scenarios. Some of them are in preliminary stages (e.g., iShare, Fiware), while others have already established communities, demonstrate running use cases, and offer a set of technical components to create a data ecosystem.

In this paper, we compare data ecosystems that, (i) have appeared in the bibliography, (ii) have already established communities, and (iii) have shown demonstrated use cases and practical applications for building data ecosystems. Thus, we considered IDS,[1] GAIA-X,[2] SOLID,[3] Data Mesh,[4] and the Ocean Protocol.[5] We present each initiative and then provide a *technical comparison* and an *applicability, outreach, and business analysis* which goes beyond the technical level, and touches on the extensibility, sustainability, and potential collaboration among data ecosystems. Throughout the rest of this article, we employ the term "*data ecosystem*", which encompasses a broader scope than the term "*data space*", as it includes not only the technical features but also the interconnected socio-economic aspects.

We derive insights into the fast-growing domain of data ecosystems, which is still relatively new and unexplored, and identify open challenges for future research and possible collaboration opportunities among different initiatives. Knowing the characteristics and differences of data ecosystems, (i) researchers can investigate relevant unexplored research topics; (ii) stakeholders can determine the data ecosystem that best serves their needs and fits their use cases; (iii) developers can leverage the results to combine modules and functionalities from different data ecosystems; and (iv) participants, e.g., councils and governments, can improve their services to provide new ones for supporting uncovered aspects. This study is intended for readers who are already somewhat familiar with the concept of data ecosystems and seek a better understanding and a comparative analysis of the existing solutions. Rather than serving as an introductory or educational resource, it provides a comprehensive analysis of the strengths and limitations of multiple data ecosystems, enabling informed decision-making.

The paper is organized as follows: Section 2 describes the state-of-the-art in the domain of data ecosystems, whereas Section 3 describes our methodology for conducting the present study. Section 4 includes an overall presentation of the emerging initiatives that have been proposed to build data ecosystems. In Section 5, we examine the selected initiatives and provide a detailed comparison of technical and non-technical aspects. Finally, Section 6 presents a discussion informed by our study, accompanied by recommendations and implications for research and practice while Section 7 outlines our conclusions and some open topics for further research.

2. State of the art

In this section, we provide an overview of the state-of-the-art, encompassing theoretical foundations (Section 2.1), and emerging initiatives for building data ecosystems (Section 2.2).

2.1. Theory and research

The "ecosystems" metaphor has been used to describe multiple and varying interrelationships between many actors and infrastructure that contribute to a resource (e.g., business, service, or software) [18]. A *data ecosystem* is composed of complex networks of organizations and individuals that exchange and use data as the primary resource [7]. Several definitions can be found for data ecosystems in the bibliography. After reviewing several scholars, [18] define a Data Ecosystem as "*a set of networks composed by autonomous actors that directly or indirectly consume, produce, or provide data and other related resources. Each actor performs one or more roles and is connected to other actors through relationships, in such a way that actors' collaboration and competition promote Data Ecosystem self-regulation*" [18]. Thus, a data ecosystem is a decentralized socio-technical network of autonomous actors (e.g., people, organizations, systems) that interactively share, manage, and use data. Their interaction is supported by roles, shared governance principles (e.g., data sovereignty, interoperability, trust, compliance), and technical architectures that enable secure and meaningful data exchanges. The fundamental purpose of data ecosystems is to facilitate trustworthy data sharing. Through collaboration, they allow actors to maintain control of their data while promoting interoperability, legal compliance, and trust. This fosters data-driven innovation and supports the creation of beneficial, domain-oriented applications [10]

[1] https://internationaldataspaces.org/
[2] https://www.gaia-x.eu/
[3] https://solidproject.org/
[4] https://www.datamesh-architecture.com/
[5] https://oceanprotocol.com/

What sets data ecosystems apart from traditional data-sharing models — such as centralized data platforms, data warehouses, and data lakes — is their inherently socio-technical and decentralized nature. Traditional models typically centralize control, governance, and storage within a single entity or infrastructure, limiting flexibility and scalability, and often resulting in fragmented trust among participants. In contrast, data ecosystems are characterized by decentralized governance, distributed storage, and dynamic interactions among autonomous participants, enabling greater scalability, flexibility, and resilience. Thus, data ecosystems uniquely integrate technological solutions with socio-economic governance frameworks to enable sustainable, trusted, and scalable data-sharing networks [12].

A taxonomy for data ecosystems demonstrating their basic key dimensions and characteristics was presented in [19]. This study shows that data ecosystems have common characteristics classified into three meta-dimensions: economic, technical, and governance. The economic dimension refers to characteristics from a business-model perspective (e.g., the applied domain and the purpose). The technical dimension refers to characteristics concerning the reference architecture of the data ecosystem (e.g., infrastructure for data sharing). The governance dimension refers to the interdependence of the data ecosystem's actors and their control over the data ecosystem's resources. In our analysis, we keep the technical dimension to cover all implementation aspects and the business dimension to cover both economic features (i.e., funding, business models) and governance aspects from the managerial point of view (i.e., partnership models).

S Oliveira et al. [5] present a systematic study on the evolution of research in data ecosystems. The study reviews relevant venues, authors, scholars, contribution types (e.g., tool, method, analysis), themes, and topics. The study emphasizes the technical knowledge and resources required to maintain a data ecosystem, the complexity of the involved tasks, the lack of actor participation and organizational structure, and privacy-related aspects. This work also reveals the need for more research in terms of both theory (e.g., introducing well-accepted definitions, conceptual models, etc.) and practice (e.g., engineering methods to improve the data ecosystem's processes, etc.). These findings serve as a starting point to conduct our comparison of the different technical and business aspects of data ecosystems.

Otto et al. [3] elucidate the transformation of data from an outcome of processes to a strategic resource for fostering data-driven innovation scenarios. In this work, basic challenges in exploiting the potential of data ecosystems are presented: trust, data sovereignty, interoperability, data governance, and compliance with legislation. We treat these challenges as fundamental principles that should be embraced by the data ecosystems to benefit all participants through robust data sharing procedures. Therefore, we examine how these challenges are currently implemented in practice over the examined data ecosystems.

Finally, [20] presented a systematic literature review of data ecosystems including *innovating* novel products and services, *engineering* systems by leveraging technology and information, and *collaborating* among multiple actors within a data ecosystem and its key themes. The identified research streams and corresponding themes have been considered in our paper by combining the practical side of the examined initiatives to extract knowledge for data ecosystems.

Existing research highlights the dynamic and complex nature of the data ecosystems domain, characterized by numerous challenges. Therefore, in this study, we aim to scrutinize existing data ecosystems and initiatives to pinpoint potential solutions and identify areas that warrant further research.

2.2. Practice and implementation

Cappiello et al. [9] provide a set of challenges for the future development of data ecosystems along with already established use cases. This work presents the practical side of implementing data ecosystems (e.g., IDS, GAIA-X). The authors identified several challenges that need to be addressed when developing a data ecosystem. These challenges include technical aspects such as building trust between participants, enabling interoperability, and non-technical challenges, such as finding the right number of participants, building a "closed" community, and agreement on legal measures.

Data spaces form an integral part of data ecosystems, providing key components for data sharing and supporting ecosystem objectives [12]. As a result, many initiatives and technologies are shared between data ecosystems and data spaces. For example, [12] present early use cases leveraging GAIA-X and IDS. However, data ecosystems can also exist without necessarily building data spaces [21]. In September 2021, the Big Data Value Association,[6] FIWARE Foundation,[7] Gaia-X, and the IDS Association formed the Data Spaces Business Alliance [22] to define a common reference technology framework, based on the technical convergence of existing architectures and models. However, certain aspects need further technical clarification, such as the lack of consensus when building a Basic Information Model. Similarly, a new initiative, the OpenDEI project[8] investigates a conceptual overview of data ecosystems to guide their development. The recently established Data Spaces Support Centre[9] offers, through its Blueprint,[10] core building blocks — reference architecture and specifications — to enable interoperable data sharing via common schemas, protocols, and agreements (the so-called Data Spaces Protocol[11]). These support the development of data spaces within ecosystems by leveraging diverse technologies. Our study identifies common technological ground and potential limitations of the examined data ecosystems.

[6] https://www.bdva.eu/
[7] https://www.fiware.org/
[8] https://www.opendei.eu/
[9] https://dssc.eu
[10] https://dssc.eu/space/BVE2
[11] https://internationaldataspaces.org/offers/dataspace-protocol/

Table 1
The examined data ecosystems' versions and related docs.

Data ecosystem initiative	Version (Arch)	Docs repository
IDS	IDS RAM V4.0 Stable	https://internationaldataspaces.org/publications/most-important-documents/
GAIA-X	GAIA-X Arch. Doc. 24.04	https://docs.gaia-x.eu/framework/
SOLID	Tech Reports 2024-06-05	https://solidproject.org/TR/
Data Mesh	Media Sources Library – Last check: 2025-03-11	https://datameshlearning.com/library/#sources
Ocean Protocol	Ocean Protocol– Last check: 2025-03-11	https://docs.oceanprotocol.com/

Curry et al. [10] focuses on ongoing efforts, including the development of Industrial Data Spaces, primarily designed for B2B data sharing, Personal Data Spaces, tailored for B2C data sharing, and Common European Data Spaces. The latter is envisioned to support various data sharing scenarios, encompassing both B2B and B2C interactions, aligning with the objectives laid out in the European Strategy for Data [2]. From these studies, we derive various data sharing scenarios, including B2B and B2C, which we then explore in our comparison.

This paper provides an overview of cutting-edge data ecosystems, with an emphasis on their technical and business facets. We leverage established theoretical principles, such as role adaptations, and technical attributes sourced from existing literature to formulate a conceptual model. This model serves as the foundation for our comparative analysis of the examined initiatives. To the best of our knowledge, there is no study which directly compares emerging data ecosystems' initiatives. The existing literature either introduces an initiative or concentrates on data ecosystems without delving into their specific offerings.

3. Methodology

In this section, we present the methodological details, including our goal, setup, and study limitations. We also introduce our proposed conceptual model, which serves as the foundation for our analysis throughout the paper, as well as our meta-architecture, which enables a more effective comparison of each ecosystem.

3.1. Data ecosystems selection

To select the initiatives for our analysis, we conducted a extensive search across academic databases including Scopus, Web of Science, and Google Scholar. We used specific keywords such as "data ecosystems", "data spaces", "decentralized data sharing", and "data economy" to identify relevant publications for data ecosystems. After identifying initiatives from the retrieved publications, we further investigated their status and maturity by exploring their respective websites and reviewing associated resources like technical documents and white papers (see Table 1). This allowed us to assess factors like progress, community size, and updates. In addition to academic sources and initiative websites, we extended our search to include the broader web using the same relevant keywords. This ensured that we did not overlook any initiatives that might not have been covered in academic documents, but were active in the field nonetheless.

We intentionally omitted certain initiatives because, at their current stage of development, they do not fully encompass all aspects of data sharing. As such, they merely provide some components or utilize technologies derived from the chosen data ecosystem initiatives. For instance, Fiware provides a suite of APIs and corresponding software blocks to facilitate data sharing. Interestingly, these could be used in conjunction with IDS to establish data spaces [23]. Similarly, iShare focuses on building a trust framework to facilitate the process of data sharing among participants in the data ecosystem. The primary goal of iShare is to utilize this framework alongside established initiatives (like IDS, GAIA-X, Fiware) to create a robust European trust network for B2B data sharing. Last, Eclipse Dataspace Components[12] offers connectors for building data ecosystems based on IDS and GAIA-X.

We also excluded from our study projects related to data platform creation (Big Data Value PPP projects) because they mainly refer to architectures and repositories of interoperable software and hardware components to enable the creation, transformation, evolution, and exploitation of data [10]. Similarly, we did not include in our study promising ongoing EU-funded projects, e.g., MobiSpaces,[13] Green.Dat.AI,[14] TANGO[15] and PPDS,[16] which have been recently launched under the DIGITAL Programme[17] or Horizon Europe Programme[18] as they are currently in the development phase.

[12] https://projects.eclipse.org/projects/technology.edc
[13] https://mobispaces.eu/
[14] https://greendatai.eu
[15] https://tango-project.eu/
[16] https://tinyurl.com/yvytpc4z
[17] https://digital-strategy.ec.europa.eu/en/activities/digital-programme
[18] HORIZON-CL4-2021-DATA-01-01, HORIZON-CL4-2021-DATA-01-03.

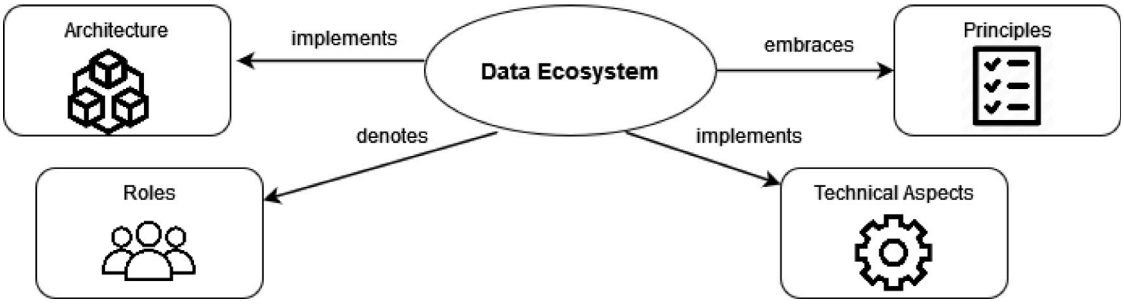

Fig. 1. Our proposed conceptual model for a data ecosystem.

3.2. Data ecosystem conceptual model

We employ a conceptual model that builds upon the following dimensions: architecture, roles, principles, and technical aspects (Fig. 1). To implement a data ecosystem, a reference *Architecture* is needed to provide the necessary components [10]. In addition, the different *Roles* for each participant should be defined to specify the functions and responsibilities of participants [24]. The roles define how different actors interact and contribute to data sharing scenarios and, as such, help in organizing and managing the collaborative aspects of the ecosystem [25]. A set of data *Principles* needs to be embraced by all participants, ensuring that data within the ecosystem is managed, shared, and utilized effectively (e.g., data sovereignty) and ethically (e.g., compliance with legislation) [3]. Finally, various technical aspects must be carefully considered to ensure the ecosystem functions effectively and efficiently concerning data utilization (e.g., data discovery and security aspects), i.e., the technical aspects are crucial for the effective and efficient functioning of the ecosystem [26]. We delve deeper into the analysis of these dimensions as outlined below (Fig. 2).

Architecture. The implementation of each data ecosystem denotes the basic components of the architecture, and defines connection methods (e.g., communication) among these components through established relationships among participants. Also in the Architecture concept, Resources are included, such as Data, Services, and Infrastructure. The implementation of the architecture depends on the ecosystem.

Roles. Roles denote a set of duties for each ecosystem's participant. Six main roles are identified [27]: (i) **data providers** make data available to other participants and provide access to data; (ii) **data brokers** facilitate interactions between data providers and users, and maintain metadata, qualities, pricing, and licenses; (iii) **service providers** offer data services (such as data analysis, certification, and data monitoring); (iv) **application developers** create functionality for using and analyzing data; (v) **infrastructure and tool providers** deliver the technical aspects and tools (e.g., user interfaces for consuming data with different kinds of devices); and (vi) **application users** consume or utilize the data. These roles can be further classified into two abstract groups: *core* and *intermediaries* [28]. **Core participants** provide the most basic functionalities, e.g., make data available. **Intermediaries** are organizations or bodies that act as agents or brokers in any aspect of the innovation process between two or more parties [29] and play peripheral roles, e.g., they facilitate the search process.

Principles. Data ecosystems need to adhere to certain principles [3]: *data sovereignty*, *data governance*, *trust*, *data interoperability*, and *compliance with legislation*. **Data sovereignty** refers to meaningful control, ownership, claims to data, and enforcement of fundamental rights of data subjects [30]. It invokes the self-determination of individuals and organizations in a data ecosystem with respect to the use of their data [31]. Data ecosystems can offer mechanisms, such as authentication and authorization, to ensure control and ownership and facilitate the self-determination of participants. **Data governance** is related to decision mechanisms to mandate responsibilities for participants as they arise from different data operations [32]. It ensures data access through specific roles, decision rights, and accountability, usually denoted through a data governance model. **Trust** is an enabler for the data economy, and as such, data ecosystems promote trustworthy data sharing, i.e., all participants need to agree on how they share their data. Trust introduces a fundamental social requirement towards building relationships among different participants within or across different data ecosystems [33]. **Data interoperability** refers to the mutual understanding in the use of data between or within data ecosystems [34] and contributes positively to the evolution of data ecosystems [33]. **Compliance with legislation** is fundamental for building trust among data ecosystem participants when sharing data, as it prevents problems that arise due to noncompliance (e.g., fines) [35]. Without adhering to these principles, ensuring the proper function of data ecosystems becomes a formidable challenge, primarily due to the absence of trust, data incompatibilities, and data anarchy [3].

Technical aspects. When implementing a data ecosystem, several technical aspects should be examined concerning data sharing. We selected the most commonly mentioned in the bibliography and tackled it through state-of-the-art initiatives. The basic information about the entities of a data ecosystem is captured in an **Information Model** to ensure data interoperability among participants [36]. **Data Storage** determines how the storage is implemented, as data may be located in different places. The storage and curation of data require reserving specific resources. Therefore, the data storage policy denotes the role of some of the participants (i.e., infrastructure and tool providers). Furthermore, the location of data storage directly affects the applicability of local privacy legislation, and, as

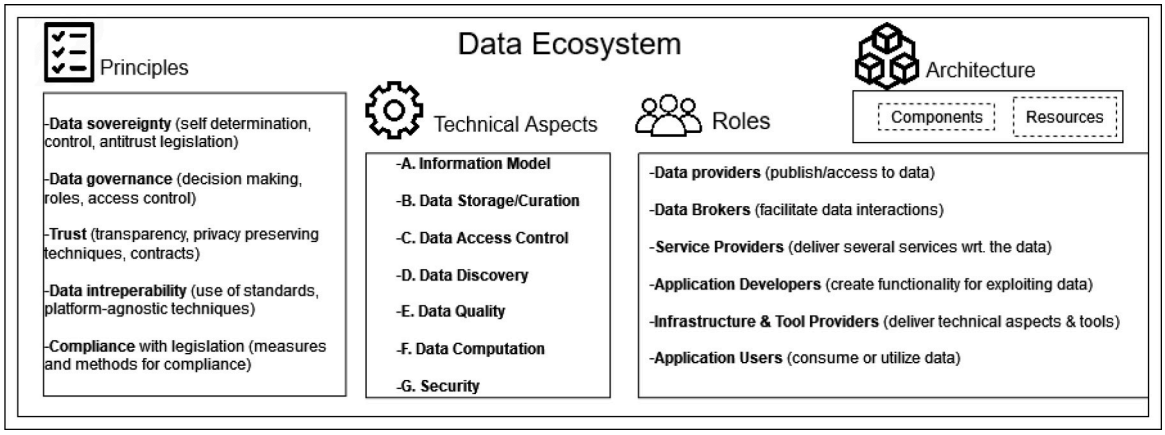

Fig. 2. Data ecosystem foundations with regards to the proposed conceptual model.

such, different rules apply to local or cloud storage [37]. **Data Access Control** authenticates and authorizes individuals to access the data they are allowed to see and use [38]. Its policy contains the rules of data sharing for all participants. These rules may be adapted to the different roles of participants to denote specific rights to access the underlying data. **Data Discovery** is the process of locating participants to enable data-driven services [39]. It facilitates the collaboration of participants, especially when a large number of participants do not know each other. **Data Computation** refers to the computational workload to perform all required transactions allowed in a data ecosystem. In a decentralized architecture, the workload can be shared among different participants with different roles [40]. **Data Quality** determines the generated value proposition for data ecosystems [3]. Mechanisms for assessing data quality are necessary because low data quality negatively impacts the data ecosystem [41]. Lastly, **Security** violations may lead to untrustworthy data sharing with negative effects in cases of data breaches, potentially for all participants of the data ecosystem [35]. The aforementioned aspects are critical for data ecosystems because they collectively form the foundation for successful data sharing and collaboration among participants.

3.3. Data ecosystem meta-architecture

The implementation of a data ecosystem involves several key parameters, each playing a crucial role in its overall functionality. To support the structured deployment of data ecosystems and enable a thorough comparative analysis, we introduce a *Meta-Architecture* (Fig. 3). This framework encompasses various layers, each designed for specific purposes and corresponding functionalities.

In particular, the *Foundational Layer* facilitates collaboration among diverse stakeholders, establishing governance mechanisms, legal compliance frameworks (e.g., GDPR, Data Act), and foundational principles such as data sovereignty and trust [3,42]. It ensures ethical and legally sound data transactions within the ecosystem. Another essential layer, referred to as the *Infrastructure Layer*, is responsible for storage, processing, and facilitating communication among system components. This layer encompasses cloud services, blockchain networks, distributed ledgers, and federated storage solutions, playing a pivotal role in defining the core architectural framework [5,42]. Furthermore, the *Services Layer* delivers fundamental capabilities, including security, access control, and information modeling. This layer integrates key technical enablers such as identity and access management (IAM), trust mechanisms, interoperability protocols, and secure data exchange frameworks. Additionally, it encompasses data marketplaces, ontologies, and semantic models, enhancing data discoverability and usability both within and across ecosystems [5,9]. The top layer, the *Application Layer*, connects the data space with both internal and external applications, including APIs, end-user applications and software services. This layer serves as the main interface for ecosystem participants, facilitating data analytics, visualization, and value extraction [3,9].

The proposed conceptual model and the meta-architecture are interrelated, with the former providing the theoretical foundation and the latter structuring its practical implementation. The key dimensions of the conceptual model — Architecture, Roles, Principles, and Technical Aspects — are systematically mapped to the corresponding layers of the Meta-Architecture (Table 2).

The conceptual model defines the fundamental characteristics of data ecosystems, while the Meta-Architecture operationalizes these characteristics into a structured, implementable framework. For example, governance principles (from the Conceptual Model) inform the policies in the Foundational Layer of the Meta-Architecture. Roles and responsibilities guide the distribution of components in the Application and Service Layers. Technical Aspects (data discovery, access control, security) are mapped directly to the services-level and infrastructure-level implementations. By aligning these two perspectives, the Meta-Architecture ensures that data ecosystems are designed in a way that adheres to established theoretical foundations while maintaining adaptability to various implementation scenarios.

Table 2
Conceptual model dimensions and their meta-architecture layers.

Conceptual model dimension	Meta-architecture layer	Key connection
Architecture	Infrastructure, Service Layer	Defines how core components (data storage, APIs, computation) are structured.
Roles	Application, Service Layer	Maps stakeholders (e.g. Application users, Data providers) to functional components.
Principles	Foundational Layer	Embeds governance, trust, sovereignty, and compliance at the regulatory level.
Technical aspects	Service, Infrastructure Layer	Covers data interoperability, access control, security, and computation.

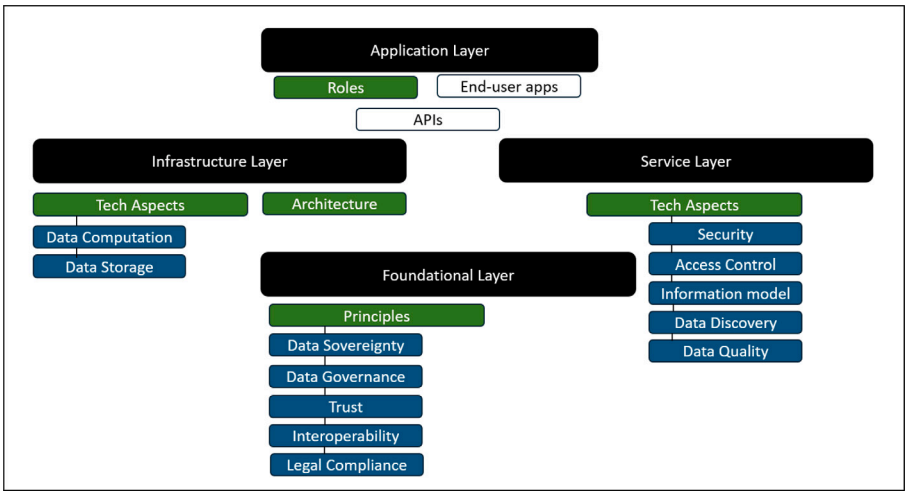

Fig. 3. The proposed meta-architecture for data ecosystems.

4. Overview of initiatives

In this section, we map the examined initiatives onto the proposed meta-architecture. IDS (Fig. 4) and GAIA-X (Fig. 6) emphasize governance and enterprise interoperability; SOLID (Fig. 8) prioritizes personal data sovereignty with decentralized storage; Data Mesh (Fig. 10) enhances enterprise data management through domain decentralization; and Ocean Protocol (Fig. 12) utilizes blockchain for decentralized marketplaces and novel economic models.

4.1. International Data Spaces (IDS)

The International Data Spaces (IDS) is built upon a common reference model, the IDS Reference Architecture Model (RAM) [28].[19,20] Since 2016, IDS has been supported by the International Data Spaces Association (IDSA), consisting of various organizations and innovators across the industry. IDS tackles the challenge of trust in cross-organizational data sharing by allowing data providers to retain control over their data (data sovereignty). It ensures that data remains at the source and is shared under strict usage policies, mitigating risks of unauthorized access or misuse. Today, IDS-based data ecosystems host several established data spaces,[21] demonstrating its practical adoption across industries.

Infrastructure Layer (Architecture and Technical Aspects). In the Infrastructure Layer, IDS[22] offers a *decentralized architecture* based on the RAM, comprising five conceptual layers (*business*, *functional*, *process*, *information*, and *system*) to support different levels of granularity. No single entity controls the network; instead, multiple entities, operated by different organizations or trusted third parties, share this role following a structured governance model as defined in the Foundational Layer. The IDS Connector is the core component, enabling secure peer-to-peer data exchange within a trusted, standardized environment. Connectors communicate via messaging services,[23] retaining usage policies set by data providers. Each Connector facilitates data exchange through exposed endpoints and operates both on-premises and in cloud environments [43].

Application Layer (Roles and applications). This layer includes the identified roles, as defined in the Business Layer of the RAM, along with the App Store, a fundamental component of the IDS architecture. The *App Store* (Fig. 5) operates as a platform for

[19] https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0
[20] https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/
[21] https://internationaldataspaces.org/adopt/data-space-radar/
[22] https://github.com/International-Data-Spaces-Association
[23] https://github.com/International-Data-Spaces-Association/IDS-Messaging-Services

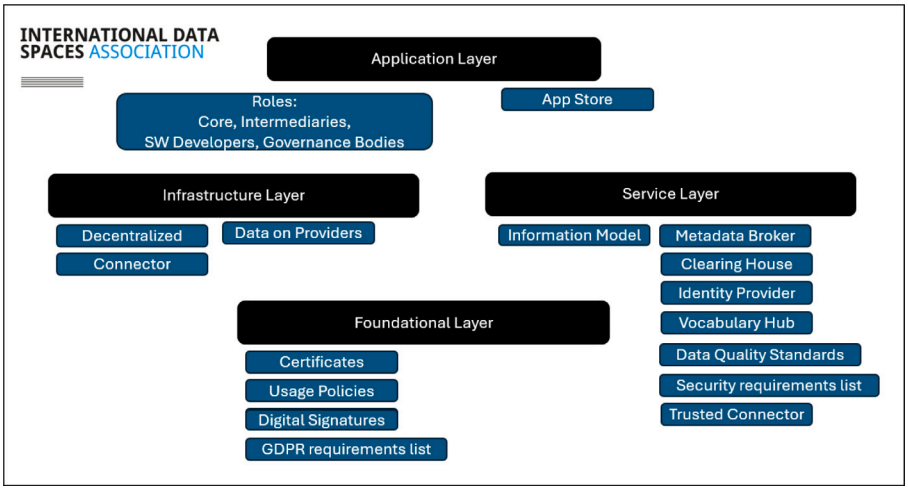

Fig. 4. An instantiation of the proposed meta-architecture within IDS data ecosystem.

distributing apps.[24] It encompasses a registry responsible for facilitating the registration, publication, and maintenance of various apps: data apps (primarily focused on data manipulation), adapter apps (geared toward facilitating access), and control apps (utilized for administrative control).

The roles in IDS fall into four main categories: *Core Participants, Intermediaries, Software Developers*, and *Governance Bodies*. Core participants include *Data Suppliers* (e.g., *Data Creators* generating IoT data, *Data Owners* with legal control, *Data Providers* making data available) and *Data Customers* (e.g., *Data Consumers* receiving data, *Service Consumers* using processed data via services, *Data Users* legally entitled users). *Intermediaries* facilitate data transactions, categorized as *Data, Service, and Vocabulary Intermediaries* (data executors, service providers, vocabulary managers), *App Stores* (app distributors), *Clearing Houses* (settlement providers), and *Identity Authorities* (identity managers). They interact via brokers: *Metadata Brokers* (for Data Intermediaries), *Service Brokers* (for Service Intermediaries), and *App Brokers* (for App Stores). *Software Developers* include *App Developers* (app creators) and *Connector Developers* (connector developers), both classified as *Application Developers*. *Governance Bodies* oversee management and certification via *Certification Bodies* and *Evaluation Facilities*, aligning with *Infrastructure & Tool Providers*. IDS RAM v4 distinguishes *typical* (business-focused) and *mandatory* (technical, ecosystem-building) roles, refining the role structure further.

Foundational Layer (Principles). Within the Meta-Architecture, IDS enforces a security-by-design approach through certificates and usage policies. Trusted connections among IDS components are established using digital signatures between connectors. In particular, *Data Sovereignty* is ensured through authentication and authorization (X.509 certificates[25]), machine-readable IDS contracts, and a security-by-design approach. *Data interoperability* is enabled by the IDS Connector(s), facilitating gateway communication for data exchange among ecosystem participants. Metadata, described using common vocabularies, enhances interoperability. *Trust* is established via data access control policies, identity management, and user certification, managed by certified Intermediaries (Identity Providers) and governance bodies (Certification Bodies). Digital signatures, known as Dynamic Attribute Tokens (DATs), validate Connectors using X.509 identity and device certificates, fostering trust among participants. To establish a trusted connection, each Connector retrieves identity information from the Identity Provider and verifies the Dynamic Attribute Provisioning Service (DAPS), which issues short-lived tokens (DATs) for secure access. This process is known as Identity Management (IM) in IDS. For *Data Governance*, IDS defines a Governance Model[26] outlining decision-making rights and processes for data usage. A responsibility assignment matrix aligns participant roles with governance roles (responsible, accountable, supporting) across all data and metadata operations. Regarding *Compliance with Legislation*, IDS provides a GDPR requirements list [44] (e.g., data encryption, anonymization, consent management) that all participants must adhere to.

Service Layer (Technical Aspects). The Service Layer in IDS includes the Information Model as a shared agreement among participants, alongside key components such as the Metadata Broker, Identity Provider, and Vocabulary Hub (Fig. 5). IDS enforces security through both technical (e.g., Trusted Connector) and non-technical measures.[27] *The Information Model* is a domain-agnostic framework structured into conceptual, declarative (IDS Vocabulary), and programmatic representations. The declarative layer, based on W3C standards like DCAT,[28] ODRL,[29] and SKOS,[30] provides a machine-readable ontology,[31] implemented in RDF Schema[32] and

[24] https://tinyurl.com/yc52c382
[25] https://www.itu.int/rec/T-REC-X.509
[26] http://tinyurl.com/yxssyddc
[27] http://tinyurl.com/23v2yyc8
[28] https://www.w3.org/TR/vocab-dcat-2/
[29] https://www.w3.org/TR/odrl-model/

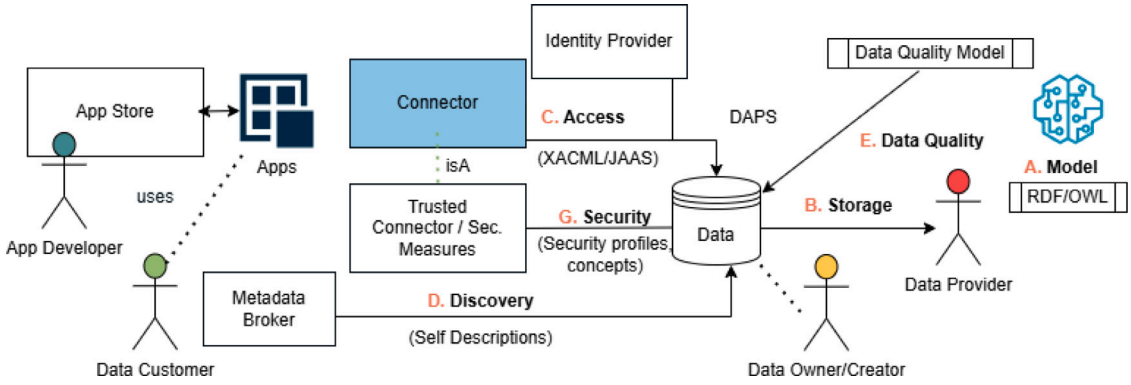

Fig. 5. The IDS data ecosystem focused on technical aspects.

OWL.[33] Core concepts include Data Assets (datasets) and Data Apps (applications). SHACL[34] and SPARQL[35] support validation and querying.

Data Access Control in IDS is managed through the Identity Provider, ensuring authentication and authorization. IDS Connectors implement authorization mechanisms using standards such as XACML[36] and JAAS,[37] with fine-grained control enforced through IDS contracts. Only certified participants gain access, ensuring trust and compliance. *Data Discovery* is facilitated by Metadata Brokers, which manage dataset self-descriptions—metadata that enables dataset search, filtering, and negotiation via IDS Connectors. *Data Quality* is ensured through evaluation processes involving Data Owners, Providers, and Brokers, following standardized metrics (model).[38] Qualification certificates validate compliance with IDS data quality requirements. *Security* in IDS is upheld by the Trusted Connector, which extends the Base Connector with multiple security profiles compliant with ISO/IEC 27001 at various levels (entry, member, central [45]). Secure communication is ensured through cryptographic methods, identity verification, and authentication.[39] Deployment options include software-based security such as TPM 2.0[40] and hardware-based security using cryptoprocessors, forming a Trusted Computing Base (TCB) [46]. IDS-RAM v4 further details security requirements and risk mitigation strategies.[41]

4.2. GAIA-X

GAIA-X[42] was first presented by the German and French Ministries of Economics in October 2019 as an initiative to enable data and services sharing. It tackles the problem of fragmentation and dependence in the cloud/data market by providing a framework to connect different data platforms and cloud services through a common federation model, without a centralized provider. GAIA-X facilitates data and services exchange across participants via Federation Services and can integrate multiple ecosystems by establishing a common operational model (GAIA-X Operational Model) based on its defined basic principles (GAIA-X Conceptual Model) [47]. This concept is supported through distinct planes: the *usage plane* for technical interoperability, the *management plane* for governance, and the *trust plane* for security via the GAIA-X trust framework. Additionally, GAIA-X[43] enables the creation of data spaces through collaboration among GAIA-X participants.

Infrastructure Layer (Architecture and Technical Aspects). The GAIA-X ecosystem within the Infrastructure Layer establishes a federated architecture, connecting multiple GAIA-X entities under a shared governance framework known as the *GAIA-X Trust Framework* [48]. In this architecture, there is no central provider; instead, interoperable cloud and edge services operate collaboratively. At the core of this reference architecture is the *GAIA-X participant*, an entity essential for domain operations with a distinct existence. Each GAIA-X participant provides *Federation Services* to address various technical aspects at the Service Layer.

30 https://www.w3.org/2004/02/skos/
31 https://w3id.org/idsa/core
32 https://www.w3.org/TR/rdf-schema/
33 https://www.w3.org/OWL/
34 https://www.w3.org/TR/shacl/
35 https://www.w3.org/TR/sparql11-query/
36 http://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-cd-03-en.html
37 https://docs.oracle.com/javase/8/docs/technotes/guides/security/jaas/JAASRefGuide.html
38 https://iso25000.com/index.php/en/iso-25000-standards/iso-25012
39 http://tinyurl.com/yeyr2kyd
40 https://trustedcomputinggroup.org/resource/tpm-library-specification/
41 http://tinyurl.com/zjpa9ya2
42 https://gaia-x.eu/
43 https://gitlab.com/gaia-x

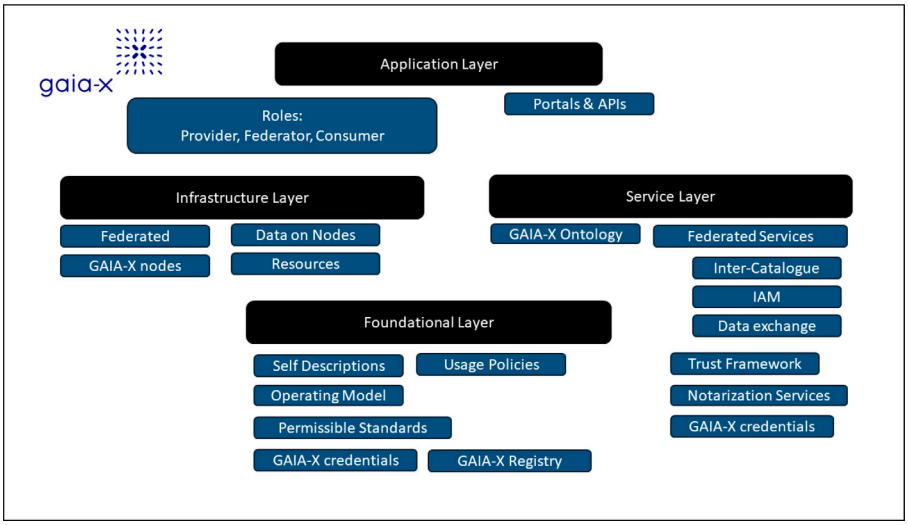

Fig. 6. An instantiation of the proposed meta-architecture within GAIA-X data ecosystem.

Additionally, a *GAIA-X node* represents a compute and storage resource. While it can store datasets, GAIA-X primarily leverages existing cloud storage infrastructure to take advantage of emerging technologies. Moreover, *Resources* describe all objects located in a GAIA-X ecosystem. They can be *Physical Resources* (e.g., a physical entity that hosts, manipulates, or interact with other physical entities), *Virtual Resources* (e.g., a dataset, configuration file, license, AI model, etc.), or *Instantiated Virtual Resources* representing an instance of a Virtual Resource (e.g., a Service Instance of a Federation Service). *GAIA-X Credentials* (formerly Self-Descriptions [49]) are core components of the GAIA-X architecture. These machine-readable files define entities within the GAIA-X conceptual model, supporting the Trust Framework and the Operating Model. They ensure validation against predefined schemas using the W3C Verifiable Credentials Data Model Standard[44] and can also serve as contracts to enforce legal agreements between GAIA-X services.

Application Layer (Roles and applications). At the Application layer, GAIA-X offers Portals and APIs that facilitate the onboarding and management of GAIA-X participants. They also support service discovery, orchestration, and provisioning of sample services, helping developers understand the technology and enabling stakeholders to adopt it effectively.

With regards to the available roles, each GAIA-X Participant may be a *Provider*, *Consumer*, or *Federator*. The Provider delivers Resources and Services in the GAIA-X ecosystem. The Consumers are the application users of the data ecosystem, who search for Service Offerings and consume Service Instances (instantiations of service offerings at runtime). The Federators are responsible for overseeing at least one *Federation Service* and function as data brokers. This is due to the fact that their catalogs are utilized to search for available participants and services. Provider and Consumer are the core roles in a business-to-business relationship, while the Federator enables their interaction by offering a set of meaningful functionalities (Fig. 7).

Foundational Layer (Principles) The Foundational Layer is built upon the Operating Model, which governs the participation of different entities [47,49,50]. This model establishes a framework to ensure trust, compliance, and interoperability while defining usage policies and legal agreements to enable data sovereignty and data governance. In this model, a "*Trust Anchor*" refers to entities recognized as trustworthy within the data ecosystem. To achieve this status, participants follow a standardized nomination process facilitated by GAIA-X labeling mechanisms, adhering to predefined GAIA-X rules that ensure transparency [51]. The GAIA-X association issues these labels, which serve as inputs for the GAIA-X compliance service, validating adherence to GAIA-X standards based on Self-Descriptions. The *GAIA-X registry* oversees this process, functioning as the validation backbone of the ecosystem.

GAIA-X enforces *Data Sovereignty* through specific policies expressed via GAIA-X Credentials (Self-Descriptions), adopting common standards such as ODRL.[45] Additionally, *Data Governance* is maintained through this transparent governance model (Operating Model), ensuring accountability and liability via well-defined Policy Rules within the GAIA-X ecosystem. The GAIA-X Association leads this model, acting as a DAO (Decentralized Autonomous Organization). Initial operational models have already been introduced, emphasizing data utilization by core GAIA-X participants and relevant data intermediaries.[46] To ensure *Trust*, the GAIA-X Trust Framework supports authentication (e.g., participant lifecycle management) and authorization mechanisms (e.g., multi-factor authentication), complemented by credential management and a decentralized identity approach. GAIA-X also introduces Trust Anchor Participants, distinguishing GAIA-X-certified participants from external participants, who have limited

[44] https://www.w3.org/TR/vc-data-model/
[45] https://www.w3.org/TR/odrl-model/
[46] https://docs.gaia-x.eu/technical-committee/architecture-document/23.10/operating_model/

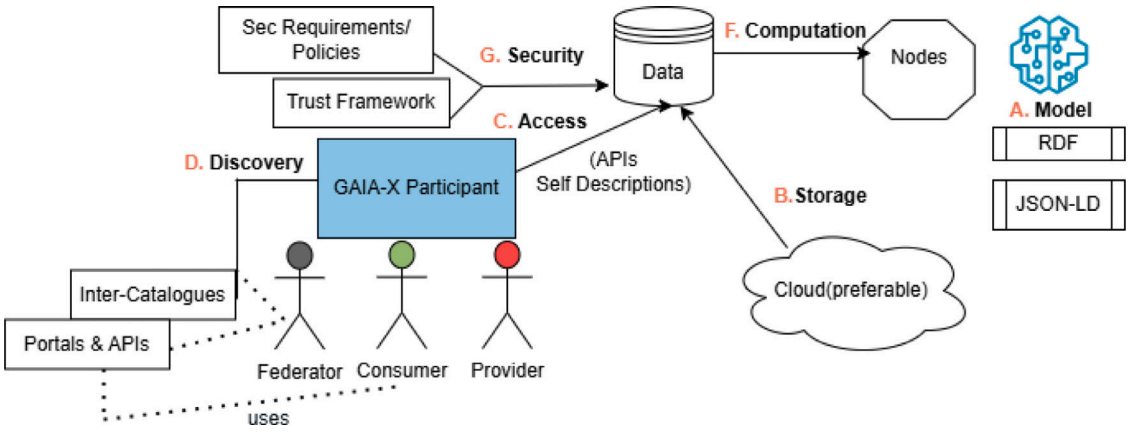

Fig. 7. The GAIA-X data ecosystem, focused on the technical aspects.

access to GAIA-X services. *Data interoperability* is facilitated through API specifications, semantic technologies, and standardized identity methods, ensuring consistent data descriptions, access control, and policy enforcement. These elements are formalized through GAIA-X Credential Schemas, validated via Shape constraints. To ensure Compliance with Established Legislation, GAIA-X relies on Contracts and Credentials. Contracts define legal agreements governing Service Instance usage, while Credentials (Self-Descriptions) document legal information about core concepts such as data ownership and policies. Moreover, GAIA-X defines *Permissible Standards*, which must be recognized by Data Protection Supervisory Authorities and comply with GDPR requirements. Compliance is measured at different Label Levels (1–3), with higher levels requiring certifications by accredited bodies. Finally, certain compliance criteria have been introduced for cloud services within the GAIA-X ecosystem.

Service Layer (Technical Aspects). The *Federation Services* practically provide the technical ground for all functionalities offered by the GAIA-X ecosystems through data and services sharing (Fig. 7). The offered Federation Services in GAIA-X follow the defined Conceptual Model.[47] However, the GAIA-X data model, still in progress, is built as an ontology. The first core model version is available online.[48] The ontology network integrates vocabularies describing key architecture elements (e.g., participants, services, and nodes). SHACL shapes define specific schemas for GAIA-X credentials.[49] The *Federation Services* are grouped into five categories: *Inter-Catalogue Synchronisation, Identity and Access Management (IAM), Data Exchange Services, Trust Framework, and Portals and APIs* [49]. *Inter-Catalogue Synchronisation* enables discovering Providers and Services. *IAM* handles authentication, authorization, and decentralized identity management. *Data Exchange Services* support sovereign data exchange via the *Data Agreement Service* (managing data-sharing agreements) and the *Data Logging Service* (verifying data transmission under *Policy Rules*). These rules ensure security (encryption, data protection, privacy) and usability (data and search policies). The *Trust Framework* enforces compliance with privacy, security, and interoperability policies. Also in the Services Layer, *Notarization Services* verify and authenticate participants' identities and data, ensuring all entities operate with validated credentials, fostering trust and integrity across the GAIA-X network.

GAIA-X enforces *Data Access Control* through Usage Policies, restricting data usage after access [52], focusing on future obligations rather than access provisions. Cryptographic signatures enable identity verification as the first step in data usage control. *Data Discovery* facilitated by Federated Catalogue services, where GAIA-X credentials identify participants and services. *Security* ensured via resource-specific policies and encryption, forming the foundation of GAIA-X compliance [51,53], while cryptographic signatures protect GAIA-X credentials to enhance security.

4.3. SOLID

Tim Berners-Lee started the SOLID project[50] in 2017, proposing a decentralized web approach to data sharing based on personal data management. SOLID addresses the problem of user data being locked into siloed applications and platforms, where it is often misused. Instead, it decouples data from applications, giving individuals full control over their own data while service providers, infrastructure, and tool providers implement SOLID using a set of standards and protocols (specifications). This is supported by an open-source implementation, the "Solid Community Server".[51]

Infrastructure Layer (Architecture and Technical Aspects). SOLID decouples data from services by allowing users to store their personal data in Data Pods (Personal Online Data stores), which can be hosted anywhere. Data pods are published anywhere

[47] https://docs.gaia-x.eu/technical-committee/architecture-document/22.10/conceptual_model/
[48] https://gaia-x.gitlab.io/gaia-x-community/gaia-x-self-descriptions/core/core.html
[49] https://gaia-x.eu/gaia-x-and-verifiable-credentials-presentations/
[50] https://github.com/solid
[51] https://github.com/CommunitySolidServer

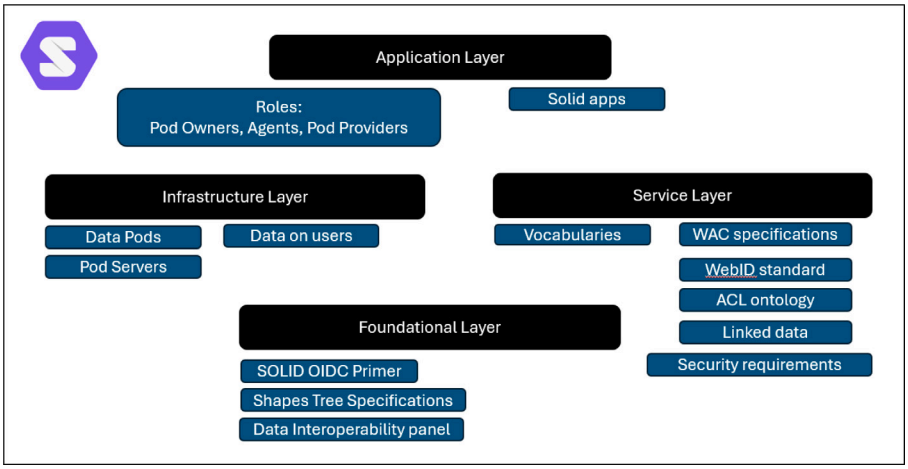

Fig. 8. An Instantiation of the proposed meta-architecture within SOLID data ecosystem.

on the web, and are stored in *Pod Servers* in a self-hosted way or in an established *Pod Provider*.[52] Inside a Data Pod, there are folders and files: the folders are mapped to machine-readable documents (Containers), and the files are mapped to documents called Resources. The actual storage is performed in Pod Servers, which are in charge of Pod Owners (application users) or Pod Providers (infrastructure providers). *Solid apps* are applications that read or write data from one or more Data Pods by applying the SOLID specification (Version 0.11.0, 2024-06-05[53]). Data Pods and apps communicate through HTTP (requests) and at a higher level through customized notifications. SOLID supports *Agents* to establish interoperability.[54] An agent is a person, social entity, or software identified by a URI.

Application Layer (Roles and applications). SOLID Apps contribute to the Application Layer by providing standardized API access to Data Pods, enabling seamless interaction with decentralized data storage. Additionally, a diverse range of open-source applications is available through the official SOLID webpage,[55] fostering the continuous development of new applications and related services. Additionally, to support the application layer, the SOLID ecosystem defines diverse roles. The owner of the Data Pod *(Pod Owner)* is the data provider in the SOLID ecosystem, who has full rights upon the data, as well as the ability to adjust the access permissions of other participants through the data access control policy. People, organizations, and applications can post a request to the public inbox of a Pod to gain data access, playing the role of application users that consume the requested data. *Pod Providers* are infrastructure providers that offer their resources to host Pod Servers. *Agents* act as data brokers to interoperate over data through the different apps.

Foundational Layer (Principles) The Foundational Layer of SOLID is based basically on two specifications, the SOLID OIDC Primer[56] which defines the basic concepts of enabling authentication and authorization among SOLID participants and the Shapes Tree Specification to define resource-oriented data structures that enable seamless data exchange.[57]

Data sovereignty is ensured, as users have full control of their data through their Data Pods. The Data Pod Owners can determine the parties to give access and declare permissions for each pod. To establish *Trust* within the SOLID ecosystem, each SOLID implementation should adopt the SOLID OIDC Primer. This specification builds upon existing web security standards, ensuring secure identity verification and controlled access through OAuth[58] for delegated access, OpenID Connect (OIDC) for identity authentication, PKCE [59] (Proof Key for Code Exchange) to prevent code interception attacks, and DPoP[60] (Demonstrating Proof-of-Possession). In terms of *Data interoperability*, SOLID provides a specification for how Agents and Applications can securely share and interact with data in a SOLID Pod.[61] It also follows Linked Data Principles, ensuring that data stored across different Data Pods remains structured and machine-readable across various applications. Additionally, it utilizes the Shapes Tree Specification to enforce a well-defined schema within SOLID Pods, making data interpretation and processing more efficient for applications. Regarding *data governance*, the

52 https://solidproject.org/users/get-a-pod
53 https://solidproject.org/TR/protocol
54 https://solid.github.io/data-interoperability-panel/specification/#agents-overview
55 https://solidproject.org//apps
56 https://solid.github.io/solid-oidc/primer/
57 https://github.com/shapetrees/specification
58 https://datatracker.ietf.org/doc/html/draft-ietf-oauth-dpop-03
59 https://www.rfc-editor.org/rfc/rfc7636
60 https://datatracker.ietf.org/doc/html/draft-ietf-oauth-dpop-03#section-4.3
61 https://github.com/solid/data-interoperability-panel

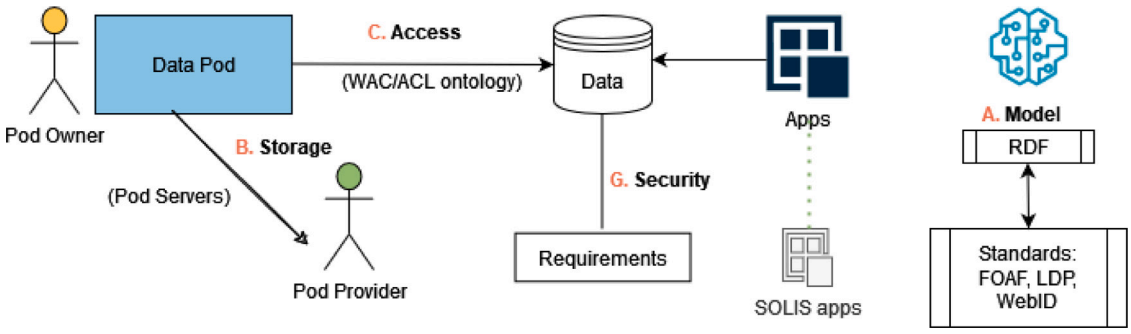

Fig. 9. The SOLID data ecosystem focused on technical aspects.

decision-making process concerning access rights primarily depends on the data access control policy. This policy allows different access modes for different operations on the data (e.g., read, write). However, there is no underlying model to regulate decisions with respect to these access modes. Regarding compliance with regulation in the SOLID specification, there are no explicit references, but some literature addresses aspects of GDPR. For instance, use cases illustrate the implementation of two GDPR rights: data portability [54] and the right of access [55]. Additionally, a proposal has been made to extend SOLID's authorization mechanism to incorporate consent management [56]. However, the lack of built-in accountability and responsibility mechanisms within legal frameworks poses challenges to ensuring full GDPR compliance [57].

Service Layer (Technical Aspects). At the Service Layer, SOLID does not provide a unified model for all information types. Instead, it leverages a set of vocabularies[62] and standards[63] to manage fundamental resource-related information. Data is structured in RDF using resources and containers, while URIs ensure proper identification. Additionally, SOLID relies on web standards and specifications to address various technical aspects at the Service Layer, ensuring interoperability and flexibility (Fig. 9).

In particular, SOLID implements a *Data Access Control* policy based on the WAC specification.[64] Also, SOLID extends the LDP[65] specification to provide a REST API for supporting various operations (read, write, append, control) on resources. Permissions are applied through RDF files included in containers that are assigned to resources. In addition, SOLID offers a set of authentication and access control libraries to be exploited when building SOLID apps. The WebID standard[66] is used for identity and authentication purposes, while the ACL ontology[67] is exploited to support authorization rules. SOLID enables Data Discovery primarily through structured metadata, and linked data. For example, through decentralized querying (e.g., via SPARQL query language for RDF data), SOLID apps can query data across Data Pods provided they have access. According to the SOLID Protocol Specification, several *Security* requirements[68] must be considered when building a SOLID ecosystem. These include enforcing TLS connections, sanitizing requests, and applying normalization and canonicalization algorithms to enhance security, data integrity, and protection against vulnerabilities.

4.4. Data Mesh

Data Mesh[69] was introduced by Zhamak Dehghani in 2019 [58] to address the challenge of scaling data analytics in large organizations. Traditional centralized data platforms often become bottlenecks; Data Mesh instead treats "data as a product" [59] and assigns each business domain responsibility for its data pipeline, from ingestion to serving consumers. Domains act as primary units responsible for creating and managing data products that directly serve consumer needs. These products can include microservices, databases, applications, and data lakes, each clearly defined and maintained by its respective domain to ensure high-quality, reliable data sharing [60].

Infrastructure Layer (Architecture and Technical Aspects). At the Infrastructure Layer, Data Mesh supports its Infrastructure (Infra) that simplifies the creation, deployment, discovery, and management of data products. Infrastructure Teams provide platforms, tooling, and frameworks, allowing domain data teams to autonomously create data products. Data Mesh supports polyglot datasets to enable several data storage policies using different technologies such as cloud services (e.g., data lakes, warehouses, or lakehouses) and processing capabilities (e.g., Spark, Kafka, Flink) for real-time and batch data.

[62] https://github.com/solid/vocab
[63] https://github.com/solid/solid#standards-used
[64] https://github.com/solid/web-access-control-spec
[65] https://www.w3.org/TR/ldp/
[66] https://www.w3.org/2005/Incubator/webid/spec/identity/
[67] http://www.w3.org/ns/auth/acl
[68] https://solidproject.org/TR/protocol#security-considerations
[69] https://datameshlearning.com

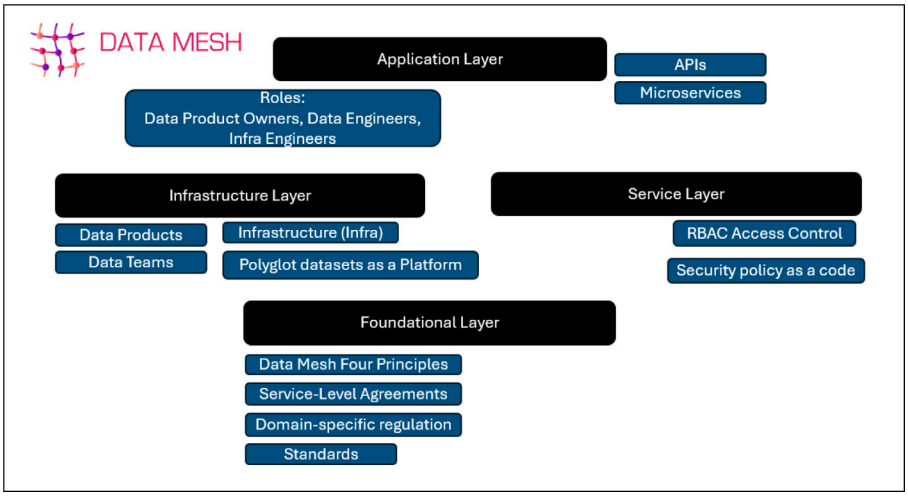

Fig. 10. An instantiation of the proposed meta-architecture within SOLID data ecosystem.

In Data Mesh, instead of having data flowing from domains into a centrally owned data lake or platform, domains host and serve their own domain datasets. This domain-oriented and decentralized architecture can be supported by exposing APIs on several decentralized data endpoints. The building blocks of a Data Mesh architecture include *Data Products*, *Data Teams*, and the offered *Infrastructure*.

Application Layer (Roles and applications). The Application Layer in Data Mesh includes the consumption of data products. This is achieved through clearly defined APIs and corresponding microservices. allowing easy access and consuming the available data. Applications could subscribe to data products via messaging brokers (e.g., Kafka) for event-driven, real-time scenarios within the Data Mesh ecosystem. Regarding the available roles, these are assigned to different *Data Teams*. Domains that provide data as a product need to be augmented with new skill sets for their team members (e.g., Data Product Owners and Data Engineers). *Data Product Owners* define the vision and roadmap for data products, take care of their consumers' satisfaction, and continuously measure and improve the quality and richness of data from their domain. This role matches with the role of data providers and data brokers, although it is wider because it gives high-level duties to data product owners (i.e., defining business-aligned KPIs). The *Data Engineers* operate internal data procedures as they occur in each domain, acting as service providers that perform several development activities (e.g., data product monitoring and versioning). *Infrastructure (Infra) Engineers* acting as infrastructure and tool providers across teams, manage data infrastructure to align domain teams with similar skills and respective responsibilities. Finally, *Data Users* are the actual data consumers, who constitute the application users for at least one data product.

Foundational Layer (Principles) The Foundation Layer in Data Mesh is based on *four principles*: *Domain-oriented ownership*: Each ecosystem component is decomposed into specific domains holding their own data ownership. Ownership shifts directly to business domain teams, bringing data closer to its source and ensuring relevant data classification and preparation by those most familiar with it. *Data as a product*: Domains are accountable for providing their data as products meeting defined quality standards, aligned with the goals of the data ecosystem. This ensures effective data management, usability, and value generation. *Self-serve data infrastructure as a platform*: Provides tools and resources enabling independent creation, deployment, discovery, and management of data products, democratizing data access and facilitating agile and scalable operations. *Federated computational governance*: Automates governance decisions through federated policies embedded within domains as code. This computational approach ensures transparency, consistency, compliance, and efficiency in governance processes.[70]

Data sovereignty, defined as self-determination over data products, is promoted through self-describing semantics encoded in data schemas. *Data governance* is supported by a federated governance model that implements global policies and operates under the supervision of a federated team in different domains. *Trust* between the data products' owners is ensured through a Service Level Objective (SLA) as an agreement around the truthfulness of data. *Data interoperability* is achieved through global standardization and specific rules to be defined for each domain to finally enable federated computational governance. To ensure *compliance with legislation*, policies can support domain-specific regulation and contractual agreements [60]. As Data Mesh pushes data ownership and accountability back to domains, additional measures should be considered at the domain level to achieve legal compliance.

Service Layer (Technical Aspects). At the Service layer, there is no predefined Data Model in Data Mesh, but a domain and conceptual model were proposed to design a Data Mesh ecosystem [61]. Regarding *Data Access Control*, a corresponding policy should be applied at the Code component. For example, the Enterprise Identity Management system (SSO) [62] and the Role Based

[70] https://tinyurl.com/ebrvyzc9

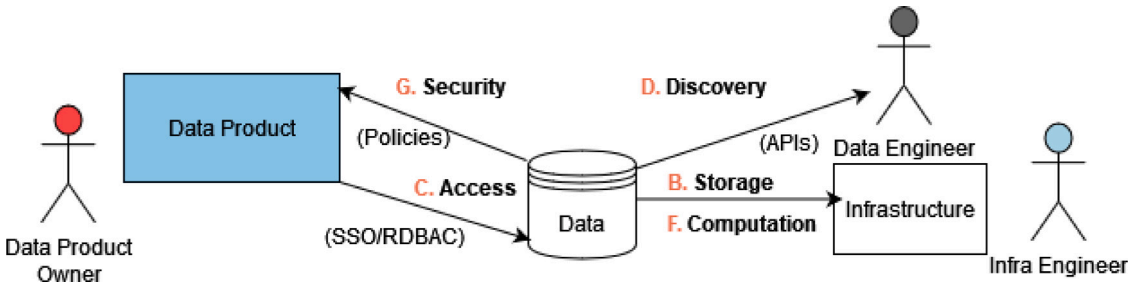

Fig. 11. The Data Mesh data ecosystem, focused on technical aspects.

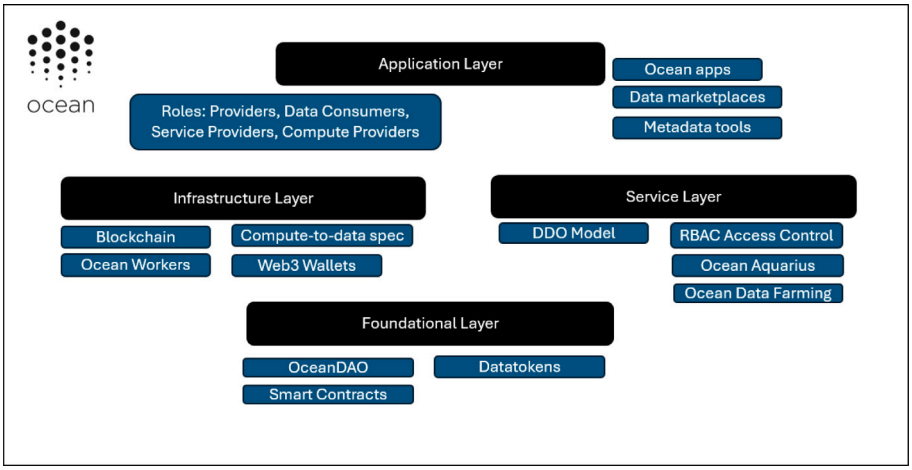

Fig. 12. An instantiation of the proposed meta-architecture within Ocean Protocol data ecosystem.

Access Control (RBAC) [63] policy definition are convenient ways to implement access control of product datasets [58]. These access control policies should change dynamically and be continuously evaluated at each point of access to the data product. To enhance *Data Discovery*, Data Mesh developers should create discoverable and understandable APIs and related documentation. Since data can be distributed across many domains, APIs should allow participants to search across available domains and data products. One proposed implementation is to have a registry of all available data products with their metadata information, such as their owners, source of origin, etc. [58]. *Data Security* is implemented through policies that define, for example, confidentiality levels for personal data and the supported type of encryption (Fig. 11). In practice, data products must follow the approach of security policy as code, i.e., security policies should be crafted in a manner that allows for version control, computational testing, enforcement, deployability, and observability.[71]

4.5. Ocean Protocol

The Ocean Protocol[72] was founded in 2017 to address data monetization and sharing in environments with low trust between parties. Traditional data marketplaces struggle with challenges related to trust, payment, and control over data usage. Ocean Protocol leverages blockchain technology, built on Ethereum,[73] to create a decentralized data marketplace where datasets are tokenized for exchange. It connects stakeholders, fosters a community, and powers data marketplaces where Ocean tokens facilitate value exchange [64]. Data owners can monetize their data while consumers gain access to diverse datasets, with blockchain tools ensuring secure, scalable transactions.

Infrastructure Layer (Architecture and Technical Aspects). The Infrastructure layer is mainly established through the formulated *blockchain* network, which is used to publish and consume data and services. The Ocean Protocol suggests a decentralized architecture where *Workers* (Providers, Publishers, Data Consumers, or Service Providers) can be connected to grow the whole data ecosystem. The Ocean Protocol ecosystem leverages Web3 wallets for decentralized data transactions, enabling users to buy,

[71] https://www.thoughtworks.com/insights/decoder/s/security-policy-as-code
[72] https://oceanprotocol.com
[73] https://ethereum.org

sell, and exchange data assets. ERC721 data NFTs represent ownership of a dataset's intellectual property, while ERC20 datatokens function as access licenses, independent of copyright ownership.[74] [75]

Data computation in Ocean Protocol allows publishers to offer compute services on their data without exposing it. The *Compute-to-Data (C2D) specification* [64] enables monetization through compute jobs while preserving data privacy. It involves Consumers (users), Operator Services (handling requests), Operator Engines (executing computations), and a Kubernetes cluster (managing workloads). This specification introduces algorithms as assets—scripts run on datasets for secure, controlled processing. Ocean Protocol does not provide data storage; publishers must store data on their servers, cloud, or decentralized storage compatible with blockchain technology. Supported options include IPFS for distributed file storage,[76] GraphQL for user and datatoken metadata,[77] and Arweave for permanent storage.[78] Simpler alternatives include smart contracts or static URLs as data sources.

Application Layer (Roles and applications). The Application Layer in Ocean Protocol features *Data Marketplaces* powered by *Ocean Apps* (or dApps in Ethereum terminology), leveraging Ocean's technology stack to enable data asset exchange. Users can discover, buy, and sell data, while applications interact with datasets to support AI and analytics development.[79]

Regarding the available roles, *Providers* offer data in the form of data assets (data providers); *Data consumers* obtain data assets for their own use (application users); and *Service providers* provide computation, storage, and algorithms for sale. *Data marketplaces* act as connectors (data brokers) between producers and consumers. They also serve as service providers, supporting users in tasks such as publishing, pricing, curation, discovery, and data consumption. *Compute Providers* play the role of infrastructure and tool providers as they provide the offering of computation on data based on the defined specification (Compute-to-Data Spec). Different roles (low-level) are applied for handling the smart contracts of the blockchain network according to different Ocean Tokens (NFT and datatokens) operations such as NFT Owner, Manager, ERC20 Deployer, Metadata Updater, and Store Updater.[80] These roles belong to the infrastructure and tool providers of our conceptual model.

Foundational Layer (Principles) In the Foundational Layer, *OceanDAO*, a Decentralized Autonomous Organization,[81] plays a vital role in community governance. It empowers OCEAN token holders to participate in decision-making by voting on project proposals. Additionally, it supports the blockchain network by funding and rewarding participants through the Network Revenue mechanism, which generates revenue using Ocean Protocol's tools, reinforcing the vision of a self-sustaining data economy. *Data sovereignty* in Ocean Protocol ensures data owners retain control while enabling sharing and monetization. They can tokenize data into data NFTs and issue *datatokens* for access, managing permissions securely. Owners decide how to grant access, including sending datatokens to OceanDAO or collaborating on new token creation. *Data interoperability* within the Ocean Protocol ecosystem is facilitated through the utilization of metadata. These metadata are expressed as *JSON* objects and facilitate the data discovery and search within data marketplaces and Ocean apps.[82] *Trust* is enabled through the Ocean *Smart Contracts*, which is a typical scenario in blockchain networks and especially in the Ethereum deployed network. These contracts ensure each datatoken is exchangeable only within the established blockchain network and its applications.[83] Smart Contracts thus enable secure data exchange and foster trust among Ocean Protocol participants.

Service Layer (Technical Aspects). The Ocean Protocol addresses all identified technical aspects in the Service Layer (Fig. 13). In particular, *the Information Model* follows the DDO specification,[84] providing a structured schema for data assets. Each asset is assigned a decentralized identifier (DID)[85] linked to a DDO file in JSON format, encapsulating metadata such as name, author, description, copyright holder, and licensing. *Data Access Control* is enforced through *Role-Based Access Control (RBAC)* at two levels. Marketplace-level permissions manage browsing, downloading, and publishing based on user roles via Ocean libraries (ocean.py, ocean.js). Asset-level permissions allow publishers to whitelist users or organizations, requiring consumers to hold a datatoken and meet DDO allow list credentials for access. *Data Discovery* is supported through *Ocean Aquarius*, which enables browsing, searching, and filtering of datasets while enhancing search efficiency and metadata accessibility via an API. *Data Quality* is incentivized through the *Ocean Data Farming* program,[86] which rewards high-quality datasets using Ocean data tokens. Data Quality is evaluated based on publisher reputation, metadata completeness, sample data availability, publisher responsiveness, and a Star Rating System for user feedback. *Security* in Ocean Protocol is ensured through Smart Contracts, encrypted URL transmission, and blockchain-specific security layers. Compute-to-Data further enhances security by enabling isolated computations, preserving privacy while allowing analysis and monetization.

74 https://docs.oceanprotocol.com/developers/architecture
75 https://docs.oceanprotocol.com/developers/contracts/datanft-and-datatoken
76 https://ipfs.tech/
77 https://github.com/oceanprotocol/ocean-subgraph
78 https://www.arweave.org/
79 https://metaschool.so/articles/build-on-ocean-protocol
80 https://docs.oceanprotocol.com/developers/contracts/roles
81 https://www.ibtimes.co.uk/ethereum-reinvents-companies-launch-dao-1557576
82 https://docs.oceanprotocol.com/developers/metadata
83 https://docs.oceanprotocol.com/developers/contracts
84 https://docs.oceanprotocol.com/developers/ddo-specification
85 https://www.w3.org/TR/did-core/
86 https://blog.oceanprotocol.com/announcing-ocean-data-farming-26c036d12f20

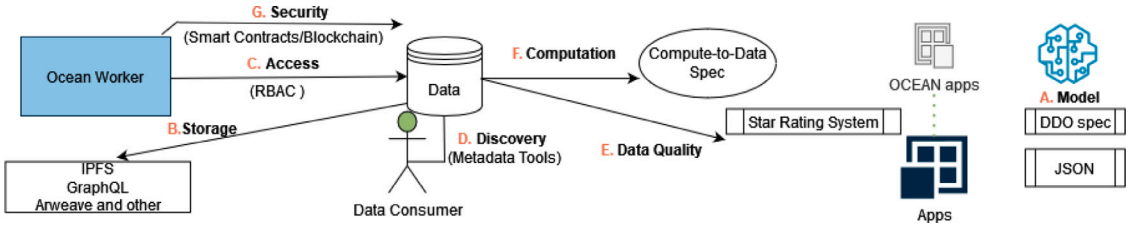

Fig. 13. The OCEAN Protocol data ecosystem, focused on technical aspects.

5. Data ecosystems review and comparison

In this section, we present a comparative analysis of each ecosystem based on its meta-architecture characteristics per layer, overall architecture, roles, principles, and technical aspects derived from our conceptual model. Additionally, we assess their applicability, outreach, and business.

5.1. Meta-architecture comparison

By mapping these five data ecosystems — IDS, GAIA-X, SOLID, Data Mesh, and Ocean Protocol — to the layers of the proposed meta-architecture, we observed similarities and differences at each layer. We summarize their merits and potential limitations in Table 3.

The shared strength across IDS, GAIA-X, and Data Mesh is their governance framework and interoperability features, precisely designed for enterprise and inter-organizational scenarios. Specifically, IDS features a Trusted Connector, Clearing House, and Metadata Broker for safe, policy-compliant data exchanges with complete governance frameworks like GDPR, digital signatures, usage policies, and governed identity management. GAIA-X also enables robust governance through its federated nodes and services that are based on a Trust Framework, notarization services, and federated catalogues and credentials for ensuring interoperability and compliance.

Conversely, SOLID and Ocean Protocol explicitly highlight decentralization. SOLID elevates higher user sovereignty of individuals at the outset, as exemplified by its data pods on the ground, WebID protocols, and access controls based on the ACL ontology, although this decentralized strategy has the twin problematics of discoverability and scalability of data. Ocean Protocol applies blockchain-based governance and marketplace protocols that can facilitate open and decentralized economic exchange, but faces regulatory and scalability concerns inherent to blockchain ecosystems.

Differences are evident when examining each layer of architecture in more detail. IDS and GAIA-X, while robust, can be susceptible to complexity and potentially increased overhead due to their extensive and elaborate governance models. SOLID's highly decentralized setup, with Pod servers and Web Access Control (WAC), provides agility but is challenged by pragmatic issues surrounding interoperability, given its individualized focus. Data Mesh, as placed, facilitates internal innovation and agility through decentralized management due to domain-driven data product owners and data engineers leveraging RBAC access management and polyglot data sets as platforms. However, this has the risk of inconsistency and fragmentation in the absence of efficient checks and balances mechanisms. Collectively, these ecosystems illustrate a spectrum from centralized, structured governance (IDS, GAIA-X) toward decentralized, user- and community-driven models (SOLID, Ocean Protocol), with Data Mesh serving as a pragmatic middle-ground solution particularly well-suited to enterprise environments.

5.2. Technical comparison

We compare the data ecosystems to: (i) assess the compatibility of different components and present state-of-the-art technologies used in their implementation; (ii) derive the merits and drawbacks of each ecosystem; (iii) identify technical challenges and potential barriers; (iv) highlight open topics for further research; and (v) seek opportunities for collaboration among the different data ecosystems.

5.2.1. Architecture and related features

All ecosystems support a decentralized architecture based on different components (Table 4), but their *degree of decentralization* varies. For example, in IDS, GAIA-X, and Data Mesh, specific components play more centralized roles in terms of decision-making processes. In IDS, the Identity Provider takes on a central role in certification processes. In GAIA-X, the registry serves as a central point for validating the operating model and cataloging tasks. Similarly, Data Teams handle governance and cataloging aspects in Data Mesh, functioning in a somewhat centralized manner. On the other hand, SOLID and Ocean Protocol offer a fully distributed approach to control. This approach offers greater flexibility for implementing ecosystems but can pose challenges when trying to establish centralized decision-making models, as there may not be a corresponding central component for this purpose.

Table 3
Comparison of data ecosystem layers in the meta-architecture.

Ecosystem	Infrastructure Layer (merits/limitations)	Application Layer (merits/limitations)	Service Layer (merits/limitations)	Foundational Layer (merits/limitations)
IDS	Secure, decentralized Connectors for controlled data sharing./Infrastructure-heavy, complexity for SMEs.	Rich application store, clear roles, secure data consumption./Certification overhead may restrict app innovation.	Strong semantic interoperability, secure Connectors./Complex semantic models, metadata overhead.	Strong governance, compliance via certifications./Complex, potentially slow certification.
GAIA-X	Federated flexible cloud and edge integration./Complex provider integration, possible service inconsistency.	Supports broad cross-industry federated applications./Limited support for consumer-facing apps.	Advanced federation services, identity, and catalog management./Federation and semantic complexity	Robust trust and governance frameworks, interoperability./Complex governance, possible bureaucratic overhead.
SOLID	Decentralized user-managed storage for privacy./Scalability challenges, reliance on users/providers	User-centric decentralized apps, enhanced privacy./Limited current adoption, scalability constraints.	Semantic interoperability, decentralized access control./No central indexing, querying may slow access	High individual sovereignty, privacy-focused./Lack of central compliance/governance.
Data Mesh	Flexible, decentralized domain-specific data architectures./Risk of infrastructure fragmentation.	Domain-driven rapid application innovation./Complex management for domain teams.	Domain-oriented APIs, computational governance./Domains' interoperability issues without strong governance.	Federated governance, agile organizational autonomy./Requires significant organizational transformation.
Ocean Protocol	Blockchain-based decentralized secure infrastructure./Blockchain scalability, cost issues.	Innovative decentralized marketplace, economic incentives./Crypto complexity, regulatory barriers.	Blockchain-driven metadata management and interoperabil-ity./Blockchain dependency limits integration.	Blockchain transparency, decentralized governance via DAO./Regulatory uncertainty, blockchain reliance.

Table 4
The architecture and basic related characteristics in data ecosystems.

Criterion	IDS	GAIA-X	SOLID	Data Mesh	Ocean Protocol
Architecture					
Basic components	Connector, Metadata Broker, App Store, Clearing House, App Store	Participants, Federation Services, GAIA-X Registry	Agent Data Pods, Data servers, Apps	Data Product (Code, Data, Metadata, Infrastructure)	Workers Apps Ocean DAO, Network
Sharing resources via					
Data	Data Suppliers, Data spaces	Participants, Data spaces	Data Pods	Data as a product	Data Marketplaces
Services	No	Yes (Federation + more)	No	No	Yes (Any kind)
Infrastructure	No	Yes (GAIA-X Nodes)	Yes (Pod Providers)	Yes (Self-serve data infrastructure)	Yes (Compute-to-Data Spec)
Data grouped by	Business	Business	Person	Domain	Marketplace

We also notice, *collaborative endeavors* in implementing the *connector*, one of the fundamental components of the IDS architecture and also involves integrating elements from GAIA-X. Specifically, the Eclipse Connector[87] is developed by the Eclipse Foundation,[88] offering an IDS-compliant connector enriched with some of the key GAIA-X concepts such as the control and data plane, as well as services like the Federated Catalogue. The EDC connector has already found utility in the Catena-X project,[89] which operates within

[87] https://github.com/eclipse-edc/Connector
[88] https://www.eclipse.org/
[89] https://catena-x.net/en/

Table 5
The comparison of roles in data ecosystem.

Data ecosystem	IDS	GAIA-X	SOLID	Data Mesh	Ocean Protocol
Core					
Data provider	Data Supplier (Data owner, data provider, data creator)	Provider	Pod owner	Data product owner	Provider (Publisher)
Application user	Data Customer	Consumer	People, Orgs, Apps	[Data User]	Data consumer
Application developer	App Developer SW Developer		[Independent devs]		[Independent devs]
Intermediary					
Data broker	Data Interm. Vocab. Interm. Metadata Broker	Federator	Agent	Data product owner	Data marketplaces
Service provider	Governance Body Service Interm Service Broker			Data engineer	Data marketplaces
Infrastructure/tool provider	Clearing House Identity Authority App Broker		Pod provider	Infrastructure engineer	Compute Provider, Low-level blockchain roles

the automotive industry. Even more, ongoing integration efforts[90] started by the Ocean Protocol community to enable trust in data sharing through the implementation of IDS connectors.

The development of apps and the use of blockchain technology are common features of all data ecosystems' architecture. The *development of apps* supported in IDS, SOLID, and the Ocean Protocol validates scenarios of data sharing in these data ecosystems. This option allows independent developers to implement and publish their applications within any of these data ecosystems (Application Developers in Table 5), contributing to ecosystem growth and sustainability. IDS is the only data ecosystem that offers different categories of apps (data, adapter, control). SOLID and Ocean Protocol further assist developers by providing the source code for apps[91] or specific templates for building apps[92] respectively.

Blockchain technology is fully exploited by the Ocean Protocol and it has started to be used by the other data ecosystems as well (e.g., IDS in the manufacturing domain[93]). Moreover, there is an ongoing effort to integrate some offerings of Ocean Protocol,[94] such as data marketplaces within the established GAIA-X data ecosystem called moveID.[95] In general, the integration between the GAIA-X participants with the nodes of the Ocean Protocol would practically lead to a richer data ecosystem, thus facilitating the offering of unified, valuable services of both ecosystems. In addition, there is an established methodology that combines pods and distributed ledgers in SOLID [65], showing potential for collaboration between SOLID and the Ocean Protocol. Consequently, blockchain technology could help combine different data ecosystems by adopting compatible technologies and offering mechanisms (such as smart contracts to enable trust) to embrace defined data ecosystem principles.

Data ecosystems can *share* a set of *resources* that are not limited to data, but can also include services and infrastructure (Table 4), amplifying the value generated by each data ecosystem that supports this kind of sharing. GAIA-X and Data Mesh support all potential options of resource sharing. Data are grouped differently according to the applied architecture and the data sharing context of each data ecosystem. However, the architecture design of each ecosystem does not prohibit alignments and changes with respect to how data are grouped to achieve collaboration among different ecosystems.

5.2.2. Roles

IDS and the Ocean Protocol offer *more fine-grained roles* than the other data ecosystems (Table 5). They provide more intermediary roles for dedicated tasks such as Vocabulary Intermediaries for supporting the role of data brokers (IDS) and Compute Providers (Ocean Protocol) for the corresponding role of infrastructure/tool provider. More fine-grained roles facilitate multi-tasking, having a positive impact on the overall performance of each data ecosystem's functionalities due to workload separation and encourage the participation of several stakeholders who might have different levels of expertise.

There are *one-to-one associations (mappings)* among the roles for the different data ecosystems. Data providers and application users are identified in all approaches. For example, Data Owner/Creator/Provider classified under the wider role of Data Supplier in IDS, Provider in GAIA-X, Pod Owner in SOLID, Data Product Owner in Data Mesh, and Provider in Ocean Protocol correspond to the same role (i.e., that of the data provider).

[90] https://github.com/oceanprotocol/provider/issues/87
[91] https://solidproject.org/apps
[92] https://oceanprotocol.com/templates
[93] https://internationaldataspaces.org/download/17278/
[94] https://github.com/deltaDAO/Ocean-Protocol-Use-Cases
[95] https://portal.moveid.eu/

Table 6
Principles comparison for data ecosystems.

Principle	IDS	GAIA-X	SOLID	Data Mesh	Ocean Protocol
Data sovereignty	Authentication, Authorization (X509 certificates), Usage Policies, Security-By-Design	Authentication via Credentials (ODRL)	Users have full control on their data	Data schemas for data products	Authorization based on fine-grained permissions, data NFTs
Data governance	Decision Making Model	Operating model	Via access rights but without an existing model	Model exists, missing details	Not Addressed
Trust	Certification (Governance Bodies), Digital signatures, IM	Trusted anchor participants, trust framework	(Specs) OpenID, OIDC Primer	Service Level Objective	Smart Contracts
Data interoperability	IDS Connector, Metadata (vocabs)	APIs, Metadata (Credentials)	Agents, Shapes spec, Metadata (RDF)	Global Standardization	Metadata (JSON)
Legal compliance	Initial lisf of GDPR requirements	Contracts, Policies (reqs), Permissible Standards	Prelim. work on consent management, GDPR rights	Considered but missing details	Not Addressed

Application developers are a vivid part of the ecosystem because they offer solutions built on top of the applied ecosystem. The IDS App Developer delivers relevant apps, while SOLID and the Ocean Protocol allow independent developers to implement their apps by exploiting existing tools and libraries. On the contrary, the role of application developers is not currently supported by GAIA-X and Data Mesh. However, topics related to, e.g., the compatibility among apps of different ecosystems and the effort estimation for connectivity are not covered yet by any data ecosystem.

Intermediary roles emerge according to each architectural component. All data ecosystems have at least the data broker role for the interaction between the data provider and the application user. Data brokers offer advanced data services through metadata (IDS), catalogues (GAIA-X), or by defining KPIs regarding data use (Data Mesh), and facilitating data operations, e.g., searching among data. *Service providers* are intermediary components supported by IDS, Data Mesh and the Ocean Protocol. Each supports different functionalities depending on its architecture and roles. IDS service providers offer SaaS services, making available a set of IDS apps (App Developer) and offering certification processes (Governance Body). In Data Mesh, service providers perform data-related tasks, e.g., data product monitoring, versioning, discovery, while service providers for the Ocean Protocol mostly sell infrastructure activities, e.g., provision of computation and storage. *Infrastructure providers* play a critical role, as they ensure the smooth operation of peripheral activities. They support different activities, e.g., data cleansing, identity provision for IDS, storing Pods in SOLID, managing data infrastructure in Data Mesh, or curating the Ocean network and handling the respective reward mechanism in the Ocean Protocol.

5.2.3. Principles

Most of our defined principles are embraced by all examined ecosystems (see Table 6). *Data Sovereignty* is ensured by default through authentication and authorization (IDS, GAIA-X, Ocean Protocol), and in SOLID due to the nature of data holding and its consumer-oriented approach. Accordingly, Data Mesh suggests different data schemas per data product that help in self-determination and control over the data.

To enable *Data Governance*, IDS presents a model that helps in the decision-making of each role and activity. The operating model in GAIA-X is slightly different from IDS, because it focuses on the compliance part of a participant with the GAIA-X ecosystem. Both ensure that new participants are compatible and compliant with the target data ecosystem. The rest of the data ecosystems do not explicitly define the data governance (Table 6); this remains an open topic for further investigation.

In all data ecosystems, the principle of *trust* is universally recognized as a necessity for the participation of every user or organization. Specifically, it is ensured through certification by specific entities in IDS (Governance Bodies) and GAIA-X (Trusted Anchor), while trusting mechanisms are used for SOLID (specifications), Data Mesh (service level objective), and Ocean Protocol (smart contracts). *Data interoperability* is based on metadata and semantic web technologies in all data ecosystems. Given the use of similar technologies, the primary challenge lies in implementing common standards through standardization efforts across data ecosystems. The establishment of global standardization schemes can further enhance data interoperability. The *compliance with legislation* principle has not been widely considered for most of the data ecosystems, besides some initial steps in the form of identifying GDPR requirements (IDS, GAIA-X) or adapting GDPR rights (SOLID). However, in addition to GDPR legislation [66], other important directives should be considered such as, the European Strategy for Data [2], the Data Governance Act [1], the Free Flow of Data Regulation (FFDR)[96] and the EU Data Act.[97]

[96] https://digital-strategy.ec.europa.eu/en/policies/non-personal-data
[97] https://ec.europa.eu/commission/presscorner/detail/en/ip_23_3491

Table 7
Technical aspects comparison for data ecosystem.

Technical aspect	IDS	GAIA-X	SOLID	Data Mesh	Ocean Protocol
Information model	RDFS, OWL	RDFS, OWL, JSON-LD	RDFS, OWL (set of vocabularies)	N/A	N/A (JSON for some tools)
Storage	Independent	Cloud (suggested)	Pod server	Independent	Independent (blockchain-compatible)
Access control	Usage control, XACML/JAAS	Usage control	WAC	RBAC/SSO	RBAC, Fine-grained permissions
Discovery	Self Descriptions by Metadata Brokers	GAIA-X Credentials by Federated catalogues	Not Addressed	APIs/Catalogues	Ocean Aquarius
Computation	Not addressed	Any node	Not addressed	Infrastructure component	Compute-to-Data Specification
Quality	Data Quality Model	Not addressed score	Not addressed	Not addressed	Data Farming Program, Star Rating System
Security	Security configurations, (Trusted Connector) P2PE, X509, Cryptographic methods, Data provenance, Identity Management	Security Reqs will be defined, Cryptographically protected Credentials	Security Considerations for HTTP requests, TLS connections etc.	Security policies as a code	Smart Contracts + SEA

5.2.4. Technical aspects

Focusing on the technical aspects (Table 7), we examine whether an *information model* is provided to capture the basic concepts and functionalities for each data ecosystem. IDS and GAIA-X provide models with an online version available. The concepts described in these models are similar (e.g., Data Assets in both IDS, while the term Data Resources is used in GAIA-X, referring to the same entity). The models of IDS and GAIA-X are compatible at the conceptual level [67]. Similarly, an initial attempt interprets the fundamental concepts of GAIA-X within the framework of Data Mesh technology[98] and reveals a notable degree of resemblance between their concepts. Nevertheless, GAIA-X relies on a service-oriented approach at the core of its conceptual model, in contrast to constraining itself solely to analytical data. In contrast to GAIA-X, the Data Mesh lacks a specific information model. SOLID uses a set of vocabularies and standards instead of a unified model, which offers flexibility but lacks consensus on how to exploit the basic terms in SOLID. Finally, Ocean protocol provides a high-level schema following the DDO standard to capture its basic entities. The exploitation of similar semantic web technologies (Table 7) on the aforementioned information models leaves room for further collaboration across data ecosystems at least at a conceptual level, which is a fundamental starting point.

Different data ecosystems support varying *data storage and curation* methods. The storage infrastructure resides at the data consumer level for IDS, in a cloud provider for GAIA-X, in a Pod Server for SOLID, or to any Ocean provider participating as data publisher. Detecting the most appropriate archiving method according to the data ecosystem (considering the available resources each time) and determining how this method affects their interoperability are still open challenges.

Different methods are used as well for *access control* (Table 7). IDS, GAIA-X, and the Ocean Protocol use fine-grained methods: IDS and GAIA-X support IDS Contracts and Credentials respectively, whereas the Ocean Protocol uses Smart Contracts. Data Mesh offers abstraction through the use of roles (RBAC), while SOLID offers a limited set of operations upon the data per resource through Web Access Control (WAC). To establish collaboration among data ecosystems, a hybrid access control mechanism is needed to combine traditional access control methods (e.g., RBAC) with more recent ones (e.g., data usage control).

Data discovery is supported through the exploitation of GAIA-X Credentials by Metadata Brokers in IDS and by Catalogues, Portals and APIs in GAIA-X. Moreover, Data Mesh relies on APIs and Catalogues to facilitate data discovery, whereas Ocean Protocol developed dedicated tools for this purpose, like Ocean Aquarius. Given the diverse data discovery methods, the collaboration among different ecosystems requires establishing a common discovery approach for the data providers, e.g., a common vocabulary such as DCAT, which is widely adopted to enable interoperability and data discovery among open data portals [68].

The aspect of *computation* has not been addressed by IDS and SOLID, while, for the rest, there are some minimum guidelines on how computation can be addressed with the guidance of infrastructure providers. This has the potential to bolster the effectiveness of machine learning and artificial intelligence techniques, thereby maximizing the utilization of the existing data [69].

Similarly, although *data quality* is very important, currently only two data ecosystems consider this aspect: IDS and Ocean Protocol. IDS suggests the creation of a Data Quality Model that could be based on a Data Quality Score. The Ocean Protocol suggests the application of rewarding schemes through the Ocean Data Farming Program and presents a Star Rating System for ranking the quality of the datasets.

IDS presents the most complete approach in terms of provenance tracking and *security* as it combines different configurations for its basic component (Connector) and considers other security requirements such as encryption, cryptographic methods, and provenance tracking. Different configurations ensure different levels of security according to the IDS use cases.

[98] https://tinyurl.com/4vtj9p7r

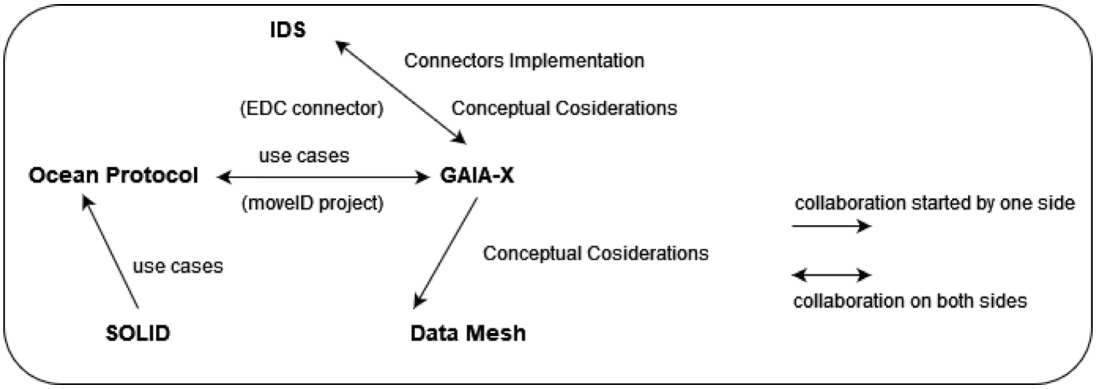

Fig. 14. Data ecosystems current collaboration efforts.

There is still room for collaboration between the different ecosystems. First steps have been made in this direction, but further steps need to be taken in the future. In Fig. 14 we summarize the collaborative efforts of all data ecosystems that have been extracted by the present study.

5.2.5. A mobility scenario for data ecosystems

To better clarify the unique technical features of each data ecosystem, we consider a mobility scenario. In IDS, participants (e.g., municipalities, transport companies) use secure connectors to establish trust through contractual data sharing policies, with centralized metadata brokers facilitating data discovery. In GAIA-X, participants leverage federation services and credentials to establish trust, making data discoverable through federated catalogs and enforcing compliance via labeled self-descriptions. SOLID, on the other hand, emphasizes individual data sovereignty by allowing users (citizens, drivers) full control over personal data stored in decentralized data Pods, with access granted explicitly through policies. Data Mesh decentralizes the responsibility to individual business domains (transport operators, municipalities), making data products discoverable via domain registries, governed by federated computational policies tailored by each domain. Finally, Ocean Protocol utilizes blockchain technologies to tokenize data assets, offering decentralized data discovery, monetization, and trusted transactions facilitated by smart contracts, ideal for complex marketplace scenarios involving diverse actors.

5.3. Applicability, outreach and business comparison

After comparing technically the data ecosystems, we examine other, more qualitative aspects that can help us understand their applicability, outreach and business orientation.

5.3.1. Applicability

IDS and GAIA-X provide the most opportunities in terms of *data sharing* because they can share any kind of data. On the contrary, SOLID can be used for any kind of data but initially focuses on personal data, Data Mesh focuses on domain data, whereas the Ocean Protocol focuses on data marketplaces (tradeable data). The majority of approaches (IDS, GAIA-X, Data Mesh) fit well with the Business-to-Business (B2B) scenario, which practically enables the cross-business involvement (i.e., data exchange), while the Ocean Protocol promotes platform-business scenarios (P2B) within the underlying blockchain network. Instead, SOLID follows the Business-to-Customer (B2C) scenario: companies develop applications, and application users are the consumers of these applications. Undoubtedly, the data sharing context of the different data ecosystems might be changed with appropriate adaptations, e.g., from B2B to B2G, by establishing agreements with government/organizations that participate in IDS, GAIA-X, and Data Mesh. Additionally, a B2B environment can be supported for SOLID if businesses develop their own SOLID apps and in the Ocean Protocol, if businesses establish their own data marketplaces.

Documentation is an important aspect that contributes to the applicability of software systems and data ecosystems as well [70]. All data ecosystems provide open-source repositories and a collection of documents for educational/training purposes, except for Data Mesh, that provides a set of blog posts and a book [60]. In addition, *open-source repositories* of code examples and relevant sources play a vital role in the continued advancement and long-term viability of data ecosystems. All data ecosystems, with the exception of Data Mesh, offer public access to these repositories. However, most data ecosystems do not provide a *starter kit*. Only IDS has recently provided a setup of open-source IDS components to implement a minimum viable data sharing solution in an IDS-based data ecosystem. [99] Ocean Protocol has several *installation requirements* that involve setting up the fundamental infrastructure, configuring it to suit the relevant operating system, and establishing the core functions of Ocean tokens. All ecosystems, except for

[99] https://github.com/International-Data-Spaces-Association/IDS-testbed

Table 8
Outreach comparison for data ecosystems.

Criterion	IDS	GAIA-X	SOLID	DataMesh	OceanProtocol
Establishment	2016	2019	2017	2019	2017
Funding scheme	NPO	NPO	Various Channels via ODI,[a] Solid companies: (e.g., Inrupt[b])	N/A	Various channels (seeds funding[c], NPO)
Funding members	Orgs (140+)[d]. 4 Communities[e]	Orgs (300+)[f]. GAIA-X community[g]	SOLID Team[h], SOLID community[i]	Slack Members (8500+), Data Mesh Community[j]	Ocean DAO[k], 24 Partners, Ocean Community[l]
Applied domains (use cases+projects)	Finance, Manufacturing, Telecom, Supply Chain, Automotive +more	Finance, Health Energy, Mobility, Agriculture +more	Finance, Health, Reviews Social Media, Games +more	Finance, Movies, Digital Media, Fashion/Retail, Software/DB	Finance, Reviews, IoT, Social Media, Agriculture +more
Events	Website announcements	Website announcements	Website announcements	Meetup group	Newsletter
Partnership model	Subscription	Open	Open/Community-driven	Open	Open/Grants
Partners role	Education, Certification, Support, Use cases.	Open-source software. Use cases.	SOLID spec. New apps, Existing apps.	Learning and start using it	Software, Dissemination (Ambassadors), Ocean DAO Grants.

[a] https://forum.solidproject.org/t/a-new-organisational-home-for-solid/8004
[b] https://inrupt.com/
[c] https://www.crunchbase.com/organization/ocean-protocol/company_financials
[d] https://internationaldataspaces.org/we/members/
[e] https://internationaldataspaces.org/make/communities/
[f] https://gaia-x.eu/who-we-are/association/
[g] https://www.gaia-x.eu/who-we-are/community
[h] https://solidproject.org/team
[i] https://solidproject.org/community
[j] https://datameshlearning.com/community-resources/
[k] https://oceanprotocol.com/explore/ecosystem
[l] https://oceanprotocol.com/explore/community/

Data Mesh, offer Docker[100] images to facilitate their installation, and all offer user support, e.g., working groups (IDS, GAIA-X), communities (IDS, Data Mesh, Ocean Protocol), forums (SOLID) and chats (Ocean Protocol).

5.3.2. Outreach and business models of the data ecosystems

IDS and GAIA-X are Non-Profit-Organizations (NPOs) that receive *funding* from their members (Table 8). SOLID has received funding from several different channels (e.g., organizations, companies, European Projects). The Ocean Protocol combines both approaches, since it has established an NPO and, in parallel, it receives funding from independent sources. The *supported members* include both organizations and businesses that participate in use cases (IDS, GAIA-X), or offer tools to build upon the applied ecosystem (SOLID), or even offer grants to teams or individuals to build their own solution on top of the data ecosystem (Ocean Protocol).

All data ecosystems exhibit wide applications in several domains by considering established use cases and projects (Table 8). Furthermore, all data ecosystems *established communities* consisting of members that contribute to the implementation (e.g., software, guidelines) or dissemination (e.g., documentation) aspects. The role of communities is fundamental for the sustainability of the ecosystems because they also help in engaging new members and keeping them active via offering them support and available resources (i.e., source, documentation, applications). Finally, all data ecosystems organize scheduled *events* to engage more participants. When considering the *partnership model*, GAIA-X, Data Mesh, and Ocean Protocol are fully open, whereas IDS requires a subscription. SOLID is community-driven, so anyone can become a partner and participate in its activities.

5.3.3. Business aspects for ecosystem participants

IDS suggests a dedicated layer, the Business Layer, to facilitate the development and use of new digital business models and the pricing concept to support different pricing models that could form the basis for creating new business models. IDS already applies the well-known subscription business model, and recently suggested more business models [71] based on the *Data Ecosystem Canvas* [72]. This canvas provides a generic model that offers guidelines on defining basic business dimensions of each data ecosystem validated by specific use cases, and it could be exploited by the other data ecosystems with appropriate adaptations. Alternative business models [73] could be examined to determine how they fit with the proposed value propositions and potential customers.

[100] https://www.docker.com/

SOLID promises disruptive innovation by transforming the role of applications and their data usage. Thus, there is a need for giant players (Data Providers), who are currently Data Controllers and Data Processors, to change their business models from data-centric to service-centric. This will lead to different business models than the existing ones. SOLID requires a mind-changing policy which envisages that the data-sharing preferences will be echoed by the new SOLID-compatible ones, and as such, they will be controlled by the SOLID application users. An alternative to creating new business models in SOLID is by adapting Web Monetization techniques, which would allow users to monetize their own data.[101] To this end, a SOLID app has been demonstrated to enable web monetization by exploiting blockchain technology for enabling payments [74].

GAIA-X recently introduced various business models, taking into account different criteria [75]. The first criterion considers the role of each participant within the GAIA-X data ecosystem. For example, Providers have the opportunity to create value propositions through data sharing, data analysis, or by providing the necessary infrastructure and tools for data acquisition and processing. The second criterion involves the primary activity within the data value chain, which includes data generation, data collection, data analysis, and data exchange. An ongoing challenge in these data models is finding ways to combine various roles and activities while defining the potential value of the data itself, as this directly impacts the applied business model.

Data Mesh has not yet specified a business model. After defining the different data domains and their responsible teams, the value proposition should be generated based on the identified needs per domain. Ocean Protocol (like SOLID) requires a disruptive way of creating a value proposition, based on exchanging data tokens. This policy requires Ocean Protocol's stakeholders to change some of their traditional procedures with respect to data sharing, and adapt these processes in a blockchain network under smart contracts. Therefore, new business models should be created to consider the context of the Ocean Protocol, either by applying blockchain-based business models [76], or by adapting existing business models accordingly: peer-to-peer, distribution-based, data licensing [73]. Currently, the Ocean Protocol does not suggest a specific business model.

6. Discussion

This section explores various options and recommendations for selecting a data ecosystem, highlighting key discussion points and their implications for research and practice based on the findings of this study.

6.1. Recommendations for choosing a data ecosystem

Selecting a data ecosystem is an intimate consideration of several trade-offs, as *no one approach fits all scenarios*. Organizations prioritizing conformity to regulatory frameworks such as GDPR and the Data Act may discover utility in formalized governance frameworks, such as in GAIA-X and IDS. These ecosystems, however, typically need significant technical expertise and infrastructural outlays that can be daunting for organizations with limited resources.

Decentralization is key to managing data, with models like SOLID and Ocean Protocol giving individuals and organizations more control over ownership of the data. These models leave people and companies in command of their data, building trust and openness. However, the lack of centralized control can create fragmentation, and interoperability and governance become more difficult to manage.

For companies that need seamless data exchange between diverse stakeholders, GAIA-X and IDS offer robust interoperability solutions. Both platforms support *structured data exchange* and mandate adherence to standards. In contrast, Data Mesh accommodates higher flexibility and domain-based solutioning, allowing companies to tailor governance models according to their business needs. While this provides advantages, it also carries risks regarding standardization and interdomain compatibility.

Economic considerations also enter the picture when selecting ecosystems. Where monetization of data is a top priority, Ocean Protocol's blockchain-based marketplace model provides an incentive-based solution. On the other hand, ecosystems such as IDS and GAIA-X are more focused on trust-based sharing, where security policy and access control take center stage over monetary transactions.

Scalability and ease of adoption further distinguish the ecosystems. Data Mesh and SOLID support organic growth by enabling the onboarding of new participants without strict procedures. However, this comes at the cost of formal compliance, which is more prominent in IDS and GAIA-X. These latter ecosystems offer well-defined frameworks but may present higher adoption barriers.

To select the right data ecosystem is use case-dependent since every model has varying trade-offs. For enterprise data sharing in B2B, IDS and GAIA-X provide secure, regulated data exchange via structured governance controls that foster trust. When personal data control is paramount, SOLID ensures individuals have full ownership and control over their data. For decentralized monetization, Ocean Protocol utilizes blockchain-based tokenization to enable secure and incentivized transactions. Data Mesh suits organizations needing scalable and flexible data management, leveraging domain-focused governance and decentralized ownership. GAIA-X and IDS benefit governments and the public sector with high interoperability, regulatory compliance, and data sovereignty assurances.

Lastly, selecting a data ecosystem requires balancing compliance, scalability, governance, and economic factors. Organizations must assess their priorities and weigh control against trust, interoperability, and operational viability to determine the best fit for their needs.

[101] https://github.com/solid/webmonetization

6.2. Research and practice implications

The use of *common features and practices* leaves room for collaboration among different ecosystems. Existing data ecosystems support the development of applications, the use of blockchain technology, commonly offered roles, and the exploitation of semantic web technologies. Also, there is established collaboration at the technical (use cases, component implementation) or at the theoretical level (conceptual considerations). The collaboration at the component level within a data ecosystem or among different data ecosystems is a challenge that can multiply the advantages of data shareability [5]. This collaboration supports ecosystem development, sustainability, and growth.

However, several challenges arise from the collaboration between data ecosystem participants, such as the need for *orchestration and management models* [4]. Towards this direction, a meta-model can be used to assist in the coordination of ecosystems by defining what elements need to be monitored and how they are related [24]. Future studies need to research the development of meta-models that facilitate effortless interaction across multiple data ecosystems, with governance, architecture, and business model compatibility.

Also, a *uniform federation service* is needed across different data ecosystems. This requires at least compatibility at the information model, as well as a common discovery approach among communication components. Furthermore, it is worth exploring a minimum soft infrastructure, referring to the fundamental shared elements, as conceptualized in the OpenDEI project. This area deserves further research. By establishing a soft infrastructure for data spaces, it guarantees the existence of legal, operational, and functional agreements. When accompanied by robust technical standards, these agreements can facilitate collaboration within or across multiple data spaces and respective data ecosystems.

A prerequisite for collaborative data sharing is *data interoperability* to ensure communication and compatibility among the different components. More work is required to establish a strategy that ensures the interoperability among data ecosystems, based on well-established technologies and standards. This strategy could be based on adaptable information components and objects, as well as on an infrastructure and development space that promotes certification for relevant applications [77]. The collaboration can be extended at the level of *applications*, by providing guidelines and resources (e.g., APIs, source code) to help developers exploit features from more than one data ecosystem. New applications and digital services can be created by joining forces and technologies.

To this end, a starting point for collaboration among data ecosystems' components could be to adapt the European Interoperability Framework (EIF) recommendations[102] by exploiting the offered toolbox (EIF toolbox[103]). The EIF Toolbox contains guidance documents on the theoretical background of the framework (EIF Pillars), offers information that highlights the implementation needs of the EIF's recommendations and principles, and provides operational solutions covering the alignment and implementation aspects of developing interoperable platforms or services. Another interesting effort is the collection of data interoperability standards,[104] which is an ongoing effort by the Data Spaces Support Center. The next step is to agree on common models, principles (e.g., FAIR [78,79]) and relevant compatible technologies. Thus, researchers could focus on creating flexible yet standardized models for semantic interoperability, metadata harmonization, and secure data exchange protocols, fostering seamless collaboration across sectors and accelerating innovation in data sharing.

In addition to common features, we identified several key *differences* among the ecosystems. Data is organized based on each architecture's implementation, with ecosystems aligning better with specific data-sharing contexts (e.g., B2B vs. B2C). IDS provides fine-grained roles, distributing responsibilities among participants, an approach that could benefit other ecosystems. IDS and GAIA-X also apply policy-based usage control, offering advantages in data governance. In contrast, the lack of centralized components in SOLID and Ocean Protocol allows for greater flexibility but poses challenges for centralized governance. These differences, when contrasted with shared features, highlight opportunities for developing hybrid solutions tailored to diverse stakeholder needs.

Another significant field of study is *measuring data quality in ecosystems*. Our analysis showed that a few platforms, such as IDS and Ocean Protocol, support data quality scoring. More research needs to examine methodologies for normalized data quality scoring, possibly in combination with AI-based validation mechanisms to achieve maximum trust and usability of shared datasets. Second, research on *incentive systems* to motivate data providers to increase the quality and usability of the data they provide can guarantee data ecosystem sustainability.

Moreover, our study highlights the necessity for research into *automated compliance mechanisms* that align with evolving regulatory frameworks such as GDPR and the European Data Act. Current manual and semi-automated compliance methods are insufficient to address the complexity and dynamic nature of modern regulations. Future studies might explore how artificial intelligence and computational governance technologies can automate compliance checks, enhance transparency, and enforce data sovereignty effectively. Automation would significantly reduce the compliance burden, increasing the attractiveness and adoption of data ecosystems.

We identified *innovative aspects* concerning data sharing and usage. For example, SOLID suggests a different way of data sharing, where the full control remains at the user side, so service providers should work together with users to gain mutual benefits. Data Mesh suggests the concept of "data as a product", a different approach to data sharing based on domain data. Domain data is classified according to different criteria, such as the sources or the customer needs. In this frame, Data Mesh aspires to change the concept of data ownership to domain ownership according to the data sharing scenario. Ocean Protocol introduces the "compute-to-data specification", enabling participants to perform computations on data within a secure environment, generating outcomes like

[102] https://tinyurl.com/3nye3mm8
[103] https://tinyurl.com/53dvt5sc
[104] https://tinyurl.com/3b2wdpby

statistical analysis or AI model development. These innovative approaches redefine traditional data-sharing paradigms and open avenues for research on user-centric governance, domain-driven data management, and privacy-preserving computation, offering practical implications for designing more adaptive, secure, and collaborative data ecosystems.

The notion of *openness* [4] (e.g., open source software development, transparent policy for joining) and flexibility (e.g., use of platform-agnostic technologies) is not only important for developers who want to contribute to a data ecosystem, but also for all involved companies or organizations which aim to play a critical role in exploiting data-driven services. In any case, developers and ecosystem participants are encouraged to actively contribute to current community-driven projects and standardization processes to help shape future directions. Through their participation in open-source development and community discussions, stakeholders can inform new standards and technologies that mirror pragmatic, real-world needs. Participation is essential for the further honing of ecosystem functionalities to make them more responsive to actual needs, hence facilitating broader acceptance and more durable implementations across different application areas.

Additionally, a *collaborative business models* [80] in the context of data ecosystems may be beneficial, as they involve multiple stakeholders with diverse needs. Its exploration can lead to new business models that create value through the sharing of data and services within the evolving landscape of data ecosystems. Some relevant business models available for established data spaces are worth examining further for potential application opportunities.[105]

In practice, findings of this research offer *critical insights to practitioners* in adopting or joining data ecosystems. These insights help stakeholders make strategic decisions when selecting ecosystems that best match their organization-specific needs. For instance, organizations focused on stringent governance may prefer GAIA-X due to its policy-oriented approach, while user-centric, data sovereignty-driven organizations may benefit more from SOLID. As strengths and weaknesses vary across ecosystems, practitioners may also combine features, for example, using Ocean Protocol's blockchain-enabled marketplace with GAIA-X's compliance support, to achieve better results.

7. Conclusion and future work

In this paper, we shed light on the emerging but unexplored field of data ecosystems by comparing state-of-the-art approaches. This study is intended for readers who are already familiar with at least one data ecosystem and seek a deeper comparative understanding of others. Rather than serving as an introductory or educational resource, it provides a comprehensive analysis of the strengths and limitations of multiple data ecosystems, enabling informed decision-making. The work is relevant to a broad spectrum of stakeholders. Researchers can further investigate relevant unexplored research topics. Stakeholders can determine the data ecosystem which best serves their needs and how their use cases can be properly adapted. Developers can exploit our work to improve their services or to provide new ones to support uncovered aspects.

We also identified several topics for future work that could contribute to the growth and sustainability of data ecosystems (Section 6.2). These include the creation of new *business models* to cover the needs of various participants, the investigation of methodologies to ensure *compliance* within an ecosystem (at component level) or among different ecosystems to facilitate the collaboration and maximize the offered benefits, and the assessment of data in terms of *data quality* and *compliance with legislation*. Since the data ecosystems change frequently, a maturity model is needed, e.g., [10] inspired by Rosemann and De Bruin [81], to characterize the features and components of data ecosystems. Additionally, acknowledging the value of quantitative methods, future research could integrate multivariate analysis techniques to facilitate evaluations of ecosystem dynamics, interactions, and performance.

Moreover, our analysis reveals the potential for further advancement in the *utilization of semantics* within the context of data ecosystems. This includes the development of well-defined vocabularies that accurately describe the essential concepts of data ecosystems. By enhancing these vocabularies, we can foster a shared understanding through conceptual modeling, thereby promoting collaboration among data ecosystems and their participants. Additionally, leveraging semantics can significantly contribute to achieving the desired interoperability among data ecosystems.

As ecosystems grow, *decision and governance models* should be further examined to denote the mandates and duties of different participants, while *assessment methods* are needed to improve the effectiveness of data ecosystems. Finally, both maturity, decision, and governance models contribute to the sustainability of data ecosystems which is an area that leaves room for further research and investigation.

Moreover, *environmental sustainability* has become a critical factor, especially given the growing use of decentralized and blockchain-based structures like Ocean Protocol. These structures inherently demand high computational capacity, which relates to energy consumption and carbon footprint. Future studies must rigorously evaluate and compare the environmental sustainability of data environments, including energy efficiency, carbon footprint, and sustainable methodologies. Such insights will help stakeholders choose or design ecosystems that not only meet functional requirements but also align with broader sustainability goals [82].

Economic viability and scalability are also fundamental to long-term success and take-up of data ecosystems. Different funding models, such as token-based economies in ecosystems like Ocean Protocol and public–private partnerships like GAIA-X, play a significant role in driving take-up and ensuring ecosystem sustainability. Assessment of these economic factors, including cost-benefit analyses, investment attractiveness potential, and financially sustainable models, will allow stakeholders to ascertain not only short-term advantage but also scalability and resilience to future market shifts.

[105] https://dssc.eu/space/SK/35520539/3+Business:+Value+and+Models

Finally, the potential of *cross-ecosystem integration* is promising. As more organizations embrace the value of complementary functionalities, such as employing IDS to protect data exchange, GAIA-X to ensure interoperable federation, and Ocean Protocol for data monetization, there is promise in building merged or hybrid ecosystems. Realizing this potential will require addressing major interoperability challenges, technology integration complexities, and the harmonization of compliance and governance frameworks. Successful cross-ecosystem integration would enable stakeholders to benefit from synergistic gains and unlock value from sharing data across contexts.

CRediT authorship contribution statement

Ioannis Chrysakis: Formal analysis, Writing – review & editing, Validation, Project administration, Investigation, Visualization, Writing – original draft, Resources, Methodology, Conceptualization. **David Chaves-Fraga:** Writing – original draft, Writing – review & editing. **Giorgos Flouris:** Writing – review & editing, Writing – original draft. **Erik Mannens:** Writing – review & editing, Supervision. **Anastasia Dimou:** Conceptualization, Formal analysis, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The described research activities were partly funded by Ghent University, Flanders Innovation & Entrepreneurship (VLAIO), and the European Union. Anastasia Dimou is partially funded by the Flemish AI Research Program (FAIR) and Flanders Make. David Chaves-Fraga is funded by the Agencia Estatal de Investigación (Spain) (PID2023-149549NB-I00), the Xunta de Galicia – Consellería de Cultura, Educación, Formación Profesional e Universidades (Centro de investigación de Galicia accreditation 2024–2027 ED431G-2023/04 and Reference Competitive Group accreditation 2022–2026, ED431C 2022/19) and the European Union (European Regional Development Fund–ERDF). This work has been partially funded by the European Union under the AgriDataValue project (Grant Agreement No. 101086461).

Data availability

No data was used for the research described in the article.

References

[1] European Data Governance Act, European Commission, 2020, Retrieved 04/04/25, from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52020PC0767.
[2] A European Strategy for Data, European Commission, 2020, Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066&from=EN.
[3] B. Otto, D. Lis, J. Jürjens, J. Cirullies, S. Opriel, F. Howar, S. Meister, M. Spiekermann, H. Pettenpohl, F. Möller, Data Ecosystems—Conceptual Foundations, Constituents and Recommendations for Action, Fraunhofer ISST, 2019, pp. 0943–1624.
[4] K. Schmück, M. Sturm, O. Gassmann, Decentralized platform ecosystems for data and digital trust in industrial environments, in: Connected Business, Springer, 2021, pp. 127–136, http://dx.doi.org/10.1007/978-3-030-76897-3_7.
[5] M.I. S Oliveira, G.d.F. Barros Lima, B. Farias Lóscio, Investigations into data ecosystems: a systematic mapping study, Knowl. Inf. Syst. 61 (2) (2019) 589–630, http://dx.doi.org/10.1007/s10115-018-1323-6.
[6] D. Lee, Building an open data ecosystem: an irish experience, in: Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance, 2014, pp. 351–360, http://dx.doi.org/10.1145/2691195.2691258.
[7] A. Zuiderwijk, M. Janssen, C. Davis, Innovation with open data: Essential elements of open data ecosystems, Inf. Polity 19 (1–2) (2014) 17–33, http://dx.doi.org/10.3233/IP-140329.
[8] S. Geisler, M.-E. Vidal, C. Cappiello, B.F. Lóscio, A. Gal, M. Jarke, M. Lenzerini, P. Missier, B. Otto, E. Paja, et al., Knowledge-driven data ecosystems toward data transparency, ACM J. Data Inf. Qual. (JDIQ) 14 (1) (2021) 1–12, http://dx.doi.org/10.1145/3467022.
[9] C. Cappiello, A. Gal, M. Jarke, J. Rehof, Data ecosystems: Sovereign data exchange among organizations), in: Dagstuhl Reports, Vol. 9, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2020, http://dx.doi.org/10.4230/DagRep.9.9.66.
[10] E. Curry, S. Scerri, T. Tuikka, Data Spaces: Design, Deployment and Future Directions, Springer Nature, 2022, http://dx.doi.org/10.1007/978-3-030-98636-0_1.
[11] M. Franklin, A. Halevy, D. Maier, From databases to dataspaces: a new abstraction for information management, ACM Sigmod Rec. 34 (4) (2005) 27–33, http://dx.doi.org/10.1145/1107499.1107502.
[12] B. Otto, M. ten Hompel, S. Wrobel, Designing Data Spaces: the Ecosystem Approach to Competitive Advantage, Springer Nature, 2022, http://dx.doi.org/10.1007/978-3-030-93975-5.
[13] B. Otto, A federated infrastructure for European data spaces, Commun. ACM 65 (4) (2022) 44–45, http://dx.doi.org/10.1145/3512341.
[14] H. Drees, D.O. Kubitza, J. Lipp, S. Pretzsch, C.S. Langdon, Mobility data space–first implementation and business opportunities, in: 27th ITS World Congress, ITS World Congres, 2021.
[15] C. Doulkeridis, G. Santipantakis, N. Koutroumanis, G. Makridis, V. Koukos, G. Theodoropoulos, Y. Theodoridis, D. Kyriazis, P. Kranas, D. Burgos, R. Jimenez-Peris, M. Duarte, M. Sakr, E. Zimanyi, A. Graser, C. Heistracher, K. Torp, I. Chrysakis, T. Orphanoudakis, E. Kapassa, M. Touloupou, J. Neises, P. Petrou, S. Karagiorgou, R. Catelli, D. Messina, M.C. Compagnucci, M. Falsetta, Mobispaces: an architecture for energy-efficient data spaces for mobility data, in: IEEE Big Data 2023, 2023, http://dx.doi.org/10.1109/BigData59044.2023.10386539.

[16] C. Doulkeridis, I. Chrysakis, S. Karagiorgou, P. Kranas, G. Makridis, Y. Theodoridis, The MobiSpaces manifesto on mobility data spaces, in: Proceedings of the 4th Eclipse Security, Ai, Architecture and Modelling Conference on Data Space, 2024, pp. 66–75, http://dx.doi.org/10.1145/3685651.3685654.
[17] V. Berkhout, C. Frey, P. Hertweck, D. Nestle, M. Wickert, Energy data space, in: Designing Data Spaces, Springer, Cham, 2022, pp. 329–341, http://dx.doi.org/10.1007/978-3-030-93975-5_20.
[18] M.I.S. Oliveira, B.F. Lóscio, What is a data ecosystem? in: Proceedings of the 19th Annual International Conference on Digital Government Research, ACM, 2018, http://dx.doi.org/10.1145/3209281.3209335.
[19] J. Gelhaar, T. Groß, B. Otto, A taxonomy for data ecosystems, in: Proceedings of the 54th Hicss, 2021, pp. 6113–6122, http://dx.doi.org/10.24251/HICSS.2021.739.
[20] D. Heinz, C. Benz, M. Fassnacht, G. Satzger, Past, present and future of data ecosystems research: a systematic literature review, in: Proceedings of the Pacis 2022, 2022.
[21] S. Scheider, F. Lauf, Data sovereign humans and the information economy: Towards design principles for human centric B2c data ecosystems, in: Proceedings of the 56th Hawaii International Conference on System Sciences, 2023, pp. 3725–3734, doi: https://hdl.handle.net/10125/103087.
[22] Data Spaces Business Alliance. Unleashing the Data Economy. Technical Convergence, DSBA, 2022, Retrieved 04/04/25, from https://data-spaces-business-alliance.eu/wp-content/uploads/dlm_uploads/DSBA-Technical-Convergence.pdf.
[23] U. Ahle, J.J. Hierro, Fiware for data spaces, Des. Data Spaces (2022) 395, http://dx.doi.org/10.1007/978-3-030-93975-5_24.
[24] M.I.S. Oliveira, L.E.R. Oliveira, M.G.R. Batista, B.F. Lóscio, Towards a meta-model for data ecosystems, in: Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age, 2018, pp. 1–10, http://dx.doi.org/10.1145/3209281.3209333.
[25] G.K. Hanssen, T. Dybå, Theoretical foundations of software ecosystems, in: Iwseco@ Icsob, CEUR-WS.org, 2012, pp. 6–17.
[26] S. Martin, P. Gautier, S. Turki, A. Kotsev, Establishment of sustainable data ecosystems, in: Recommendations for the Evolution of Spatial Data Infrastructures, Vol. 30626, EUR, 2021, http://dx.doi.org/10.2760/04462.
[27] A. Immonen, M. Palviainen, E. Ovaska, Requirements of an open data based business ecosystem, IEEE Access 2 (2014) 88–103, http://dx.doi.org/10.1109/ACCESS.2014.2302872.
[28] IDSA, IDS Reference Architecture Model Version 4.0, IDSA, 2023, Retrieved 04/04/25, from https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0.
[29] J. Howells, Intermediation and the role of intermediaries in innovation, Res. Policy 35 (5) (2006) 715–728, http://dx.doi.org/10.1016/j.respol.2006.03.005.
[30] P. Hummel, M. Braun, M. Tretter, P. Dabrock, Data sovereignty: a review, Big Data Soc. 8 (1) (2021) 1–17, http://dx.doi.org/10.1177/2053951720982012.
[31] M. Jarke, B. Otto, S. Ram, Data sovereignty and data space ecosystems, Bus. Inf. Syst. Eng. 61 (5) (2019) 549–550, http://dx.doi.org/10.1007/s12599-019-00614-2.
[32] V. Khatri, C.V. Brown, Designing data governance, Commun. ACM 53 (1) (2010) 148–152, http://dx.doi.org/10.1145/1629175.1629210.
[33] J. Gelhaar, B. Otto, Challenges in the emergence of data ecosystems, in: Pacis, 2020, p. 175.
[34] C.H. Asuncion, M.J.v. Sinderen, Pragmatic interoperability: a systematic review of published definitions, in: Enterprise Architecture, Integration and Interoperability, Springer, 2010, pp. 164–175, http://dx.doi.org/10.1007/978-3-642-15509-3_15.
[35] M.S. Gal, O. Aviv, The competitive effects of the GDPR, J. Compét. Law Econ. 16 (3) (2020) 349–391, http://dx.doi.org/10.1093/joclec/nhaa012.
[36] L. Nagel, D. Lycklama, Design principles for data spaces, 2021, Retrieved 04/04/25, from https://design-principles-for-data-spaces.org/.
[37] C. Tankard, What the GDPR means for businesses, Netw. Secur. 2016 (6) (2016) 5–8, http://dx.doi.org/10.1016/S1353-4858(16)30056-3.
[38] Z. Liu, W. Gu, J. Xia, Review of access control model, J. Cyber Secur. 1 (1) (2019) 43–50, http://dx.doi.org/10.1016/j.procs.2021.03.056.
[39] E. Curry, Real-time Linked Dataspaces, Springer, 2020, http://dx.doi.org/10.1007/978-3-030-29665-0.
[40] A.J. Ferrer, J.M. Marquès, J. Jorba, Towards the decentralised cloud: Survey on approaches and challenges for mobile, ad hoc, and edge computing, ACM Comput. Surv. 51 (6) (2019) 1–36, http://dx.doi.org/10.1145/3243929.
[41] T.C. Redman, The impact of poor data quality on the typical enterprise, Commun. ACM 41 (2) (1998) 79–82, http://dx.doi.org/10.1145/269012.269025.
[42] S. Scheider, F. Lauf, F. Möller, B. Otto, A reference system architecture with data sovereignty for human-centric data ecosystems, Bus. Inf. Syst. Eng. 65 (5) (2023) 577–595, http://dx.doi.org/10.1007/s12599-023-00816-9.
[43] G. Giussani, S. Steinbuss, Data Connector Report. No. 16, IDSA, 2024, Retrieved 04/04/25, from https://zenodo.org/records/13838396.
[44] S. Steinbuss, et al., Gdpr Requirements and Recommendations for the IDS Reference Architecture Model, IDSA, 2019, Retrieved 04/04/25, from https://doi.org/10.5281/zenodo.5675903.
[45] H. Pettenpohl, M. Spiekermann, J.R. Both, International data spaces in a nutshell, in: Designing Data Spaces, Springer, Cham, Switzerland, 2022, pp. 29–40, http://dx.doi.org/10.1007/978-3-030-93975-5_3.
[46] G.H. Nibaldi, Specification of a Trusted Computing Base, Tech. Rep. M79-228, MITRE Corp., 1979.
[47] GAIA-X, GAIA-X Architecture Document 24.04, Gaia-X Association, 2024, Retrieved 04/04/25, from https://docs.gaia-x.eu/technical-committee/architecture-document/24.04.
[48] GAIA-X, GAIA-X Trust Framework 22.10, Gaia-X Association, 2022, Retrieved 04/04/25, from https://docs.gaia-x.eu/policy-rules-committee/trust-framework/22.10/.
[49] GAIA-X, GAIA-X Architecture Document 22.10, Gaia-X Association, 2022, Retrieved 04/04/25, from https://docs.gaia-x.eu/technical-committee/architecture-document/22.10/.
[50] GAIA-X, GAIA-X Architecture Document 23.10, Gaia-X Association, 2023, Retrieved 04/04/25, from https://docs.gaia-x.eu/technical-committee/architecture-document/23.10.
[51] GAIA-X, GAIA-X Compliance Document 24.11, Gaia-X Association, 2024, Retrieved 04/04/25, from https://docs.gaia-x.eu/policy-rules-committee/compliance-document/24.11/.
[52] H. Tardieu, Role of Gaia-X in the European data space ecosystem, in: Designing Data Spaces: The Ecosystem Approach To Competitive Advantage, Springer International Publishing Cham, 2022, pp. 41–59, http://dx.doi.org/10.1007/978-3-030-93975-5_4.
[53] GAIA-X, GAIA-X Policy Rules Conformity Document 23.10, Gaia-X Association, 2023, Retrieved 04/04/25, fro https://docs.gaia-x.eu/policy-rules-committee/policy-rules-conformity-document/23.10/.
[54] G. De Mulder, B. De Meester, P. Heyvaert, R. Taelman, A. Dimou, R. Verborgh, Prov4itdata: Transparent and direct transferof personal data to personal stores, in: Companion Proceedings of the Web Conference 2021, 2021, pp. 695–697, http://dx.doi.org/10.1145/3442442.3458608.
[55] E. Francesconi, et al., Automating the response to gdpr's right of access, in: Legal Knowledge and Information Systems: JURIX 2022: The Thirty-Fifth Annual Conference, Saarbrücken, Germany, 14-16 December 2022, Vol. 362, IOS Press, 2022, p. 170, http://dx.doi.org/10.3233/FAIA220462.
[56] L. Debackere, P. Colpaert, R. Taelman, R. Verborgh, A policy-oriented architecture for enforcing consent in solid, in: Proceedings of the 2nd International Workshop on Consent Management in Online Services, Networks and Things, 2022, pp. 1–9, http://dx.doi.org/10.1145/3487553.3524630.
[57] H.J. Pandit, Making sense of solid for data governance and GDPR, Inf. 14 (2) (2023) 114, http://dx.doi.org/10.3390/info14020114.
[58] Z. Dehghani, How to move beyond a monolithic data lake to a distributed data mesh, 2019, Retrieved 04/04/25, from https://martinfowler.com/articles/data-monolith-to-mesh.html.
[59] S. Narayan, Products over projects, 2018, Retrieved 04/04/25, from https://martinfowler.com/articles/products-over-projects.html.
[60] Z. Dehghani, Data Mesh, O'Reilly, 2022.
[61] I. Machado, C. Costa, M.Y. Santos, Data-driven information systems: the data mesh paradigm shift, in: Information Systems Development: Crossing Boundaries Between Development and Operations (Devops) in Information Systems, Universitat Politècnica de València, 2021.

[62] V. Kumar, A. Bhardwaj, Identity management systems: a comparative analysis, Int. J. Strateg. Decis. Sci. (IJSDS) 9 (1) (2018) 63–78, http://dx.doi.org/10.4018/IJSDS.2018010105.
[63] R.S. Sandhu, Role-based access control, in: Advances in Computers, Vol. 46, Elsevier, 1998, pp. 237–286, http://dx.doi.org/10.1016/S0065-2458(08)60206-5.
[64] T. McConaghy, Ocean protocol: Tools for the Web3 data economy, in: Handbook on Blockchain, Springer, 2022, pp. 505–539.
[65] M. Ramachandran, N. Chowdhury, A. Third, J. Domingue, K. Quick, M. Bachler, Different ways to connect solid pods and blockchain, in: Companion Proceedings of the Web Conference 2020, 2020, pp. 645–649, http://dx.doi.org/10.1145/3366424.3385759.
[66] General Data Protection Regulation, European Commission, 2016, Retrieved 04/04/25, from https://eur-lex.europa.eu/eli/reg/2016/679/oj.
[67] B. Otto, et al., GAIA-X and IDS, IDSA, 2021, Retrieved 04/04/25, from https://doi.org/10.5281/zenodo.5675897.
[68] S. Neumaier, J. Umbrich, A. Polleres, Lifting data portals to the web of data, in: Workshop on Linked Data on the Web Co-Located with 26th International World Wide Web Conference, WWW 2017, CEUR-WS.org, 2017.
[69] M.I. Jordan, T.M. Mitchell, Machine learning: Trends, perspectives, and prospects, Sci. 349 (6245) (2015) 255–260, http://dx.doi.org/10.1126/science.aaa8415.
[70] N.J. Kipyegen, W.P. Korir, Importance of software documentation, Int. J. Comput. Sci. Issues (IJCSI) 10 (5) (2013) 223.
[71] C. Mertens, et al., New Business Models for Data Spaces Grounded in Data Sovereignty, IDSA, 2021, Retrieved 04/04/25, from https://internationaldataspaces.org/download/21261/.
[72] O. Pitkänen, et al., Rulebook for a Fair Data Economy, SITRA, 2019, Retrieved 04/04/25, from https://www.sitra.fi/en/publications/rulebook-for-a-fair-data-economy/#download-the-rulebook.
[73] S. Gupta, 50 Types of Business Models (2022) – The Best Examples of Companies Using It, Business Strategy Hub, 2021, Retrieved 04/04/25, from https://bstrategyhub.com/50-types-of-business-models-the-best-examples-of-companies-using-it/.
[74] H. Becker, H. Vu, A. Katzenbach, C.H.-J. Braun, T. Käfer, Monetising resources on a solid pod using blockchain transactions, in: European Semantic Web Conference, Springer, 2021, pp. 49–53, http://dx.doi.org/10.1007/978-3-030-80418-3_9.
[75] P. Kraemer, C. Niebel, A. Reiberg, Gaia-X and Business Models, Gaia-X Association, 2023, Retrieved 04/04/25, from https://gaia-x-hub.de/wp-content/uploads/2023/11/GX-White-Paper-Business-Models.pdf.
[76] A.Y.L. Chong, E.T. Lim, X. Hua, S. Zheng, C.-W. Tan, Business on chain: a comparative case study of five blockchain-inspired business models, J. Assoc. Inf. Syst. 20 (9) (2019) 9, http://dx.doi.org/10.17705/1jais.00568.
[77] P. Pagano, L. Candela, D. Castelli, Data interoperability, Data Sci. J. 12 (2013) 19–25, http://dx.doi.org/10.2481/dsj.GRDI-004.
[78] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, et al., The fair guiding principles for scientific data management and stewardship, Sci. Data 3 (1) (2016) 1–9, http://dx.doi.org/10.1038/sdata.2016.18.
[79] M. Hauff, L.M. Comet, P. Moosmann, C. Lange, I. Chrysakis, J. Theissen-Lipp, Fairness in dataspaces: the role of semantics for data management, in: 2nd International Workshop on Semantics in Dataspaces, 2024, http://dx.doi.org/10.5281/zenodo.13120339.
[80] P. Grefen, Service-dominant business engineering with base/x: Business modeling handbook, in: BASE/X Handbooks, Vol. 1, CreateSpace Independent Publishing Platform, 2015.
[81] M. Rosemann, T. De Bruin, Application of a holistic model for determining bpm maturity, BP Trends 2 (2005) 1–21.
[82] I. Chrysakis, E. Agorogiannis, N. Tsampanaki, M. Vourtzoumis, E. Chondrodima, Y. Theodoridis, D. Mongus, B. Capper, M. Wagner, A. Sotiropoulos, F. Coelho, C. Brito, P. Protopapas, D. Brasinika, I. Fergadiotou, C. Doulkeridis, A data spaces architecture for enhancing green AI services, in: 2025 Design, Automation & Test in Europe Conference & Exhibition (DATE), IEEE, 2025, pp. 1–7, http://dx.doi.org/10.23919/DATE64628.2025.10992729.

Ioannis Chrysakis is a Joint Ph.D. Candidate in Information Engineering Technology at the University of Ghent and KU Leuven. He is member of IDLab, imec in University of Ghent and member of Declarative Languages and Artificial Intelligence Research Group (DTAI) in KU Leuven. He has co-authored more than 30 scientific peer-reviewed papers. His research interests include data spaces, data privacy, data economy, crowdsourcing and semantic web. He has been a member of the program committee and reviewer in several international journals, conferences and workshops. He has a strong record of successfully managing and delivering more than 20 European and national R&D projects across domains such as artificial intelligence, data privacy, data spaces, mobility, digital agriculture, cultural heritage, and ambient intelligence. Currently, he works as a Research and Innovation project manager in Netcompany S.A. and is affiliated as a Research Associate with the Information Systems Laboratory of FORTH-ICS.

David Chaves-Fraga is an assistant professor at Universidade de Santiago de Compostela (USC, Spain). He is also a researcher at the Center for Research in Intelligent Technologies (CiTIUS@USC), and he was previously a research collaborator at the Declarative Languages and Artificial Intelligence Research Group (DTAI) at KU Leuven, Belgium. His main research lines are focused on data management techniques for data integration systems, mainly the construction of Knowledge Graphs from (semi)structured sources using declarative mapping rules. He has been the principal researcher of the EU Public Procurement Data Space since 2022.

Giorgos Flouris is Research Director (Grade A') in FORTH-ICS. His research interests lie mainly in the broad areas of Knowledge Representation and Reasoning, (Symbolic) Artificial Intelligence, Semantic Technologies and Knowledge Graphs. Giorgos has been involved in several projects (mostly European) and published more than 150 papers in various venues. He has received awards for co-authored publications in various conferences (including STAIRS-06, ISWC-09, ISWC-15, SSWS-18, CLAR-21, SEMAPRO-22 and others). Giorgos is the coordinator of the Symbolic AI Group (SymbAI) of the Information Systems Laboratory of FORTH-ICS.

Prof. dr. ir. Erik Mannens is an independent AI expert (bridging the Knowledge gap from Research to Industry), Keynote Speaker & author of the book "Sustainable AI" (see: sustainable-ai.be), and both parttime Full Professor @ UAntwerp (Sustainable AI) and @ Ghent University (Semantic Intelligence). The last 20 years he successfully managed +160 "interdisciplinary" projects (amounting +50M euro of Funding) and managed teams of 50 to 125 researchers. He received his PhD degree in Computer Science Engineering (2011) at UGent, his Master's degree in Computer Science (1995) at K.U. Leuven & his Master's degree in Electro-Mechanical Engineering (1992) at KAHO Ghent.

Anastasia Dimou is a tenure-track assistant professor at the Declarative Languages and Artificial Intelligence Research Group (DTAI) section of Computer Science Department at KU Leuven. Anastasia conducts research on the topic of Semantic Web technologies and Knowledge Graphs, an emerging field in the spectrum of Artificial Intelligence. Her research is focused on enabling Semantic Web applications by facilitating the generation of knowledge graphs with the help of Machine Learning and fusing Semantic Web and Machine Learning technologies in general taking into consideration data privacy aspects.