



NextGEM

Next Generation Integrated Sensing and Analytical System for Monitoring and Assessing Radiofrequency Electromagnetic Field Exposure and Health

D6.7: Network provisioning and links with EU health data space – Final report

Document Summary Information

Start Date	01/07/2022	Duration	48 months
Project URL	https://www.nextgem.eu/		
Deliverable	D6.7: Network provisioning and links with EU health data space – Final report		
Work Package	WP6	Task	T6.2
Contractual due date	28/02/2025	Actual submission date	28/02/2025
Type	Report	Dissemination Level	PU-Public
Lead Beneficiary	FORTH	Deliverable Editor	Petros Zervoudakis, Panos Chatziadam (FORTH)



This project has received funding from the European Union's Horizon Europe research and innovation programme under the Grant Agreement No 101057527

Contributors and Peer Reviewers

Contributors
Petros Zervoudakis, Panos Chatziadam, Nikolaos Petroulakis, Alexandros Kornilakis, Peni Stathogiannopoulou (FORTH), Dimitris Laskaratos (ICOM), Stefanos Fafalios (SANL), Nicolas Louca (EBOS), Maryse Ledent (SC)
Peer Reviewers
Dimitris Laskaratos (ICOM), Stefanos Fafalios (SANL)

Revision history (including peer-reviewing and quality control)

Version	Issue Date	Changes	Contributor(s)
v0.1	21/11/2024	Table of Contents provided	Petros Zervoudakis, Panos Chatziadam, Nikolaos Petroulakis (FORTH)
v0.2	06/12/2024	Sections populated with the Task leaders	Petros Zervoudakis, Panos Chatziadam, Nikolaos Petroulakis (FORTH)
v0.3	20/12/2024	Section defined, assigned, and agreed	Petros Zervoudakis, Panos Chatziadam, Nikolaos Petroulakis (FORTH)
v0.4	10/01/2025	First contributions	All partners
v0.5	24/01/2025	Integration and harmonization	Petros Zervoudakis, Panos Chatziadam, Nikolaos Petroulakis (FORTH)
v0.6	03/02/2025	Second contributions and updates	All partners
v0.7	10/02/2025	Complete version ready for peer review	Panos Chatziadam (FORTH)
v0.8	17/02/2025	Peer review	Dimitris Laskaratos (ICOM), Stefanos Fafalios (SANL)
v0.9	21/02/2025	Comments addressed from peer review, technical and quality assurance	Mats-Olof Mattson (SPi), Nicolas Louca (eBOS), Panos Chatziadam (FORTH)
v1.0	28/02/2025	Final review and submission	Nikolaos Petroulakis (FORTH)

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them.”

While the information contained in the documents is believed to be accurate, the authors(s) or any other participant in the NextGEM consortium make no warranty of any kind with regard to this material including, but not limited to the implied warranties of merchantability and fitness for a particular purpose.

Neither the NextGEM Consortium nor any of its members, their officers, employees, or agents shall be responsible or liable in negligence or otherwise howsoever in respect of any inaccuracy or omission herein.

Without derogating from the generality of the foregoing neither the NextGEM Consortium nor any of its members, their officers, employees, or agents shall be liable for any direct or indirect or consequential loss or damage caused by or arising from any information advice or inaccuracy or omission herein.

Copyright message

© NextGEM Consortium. This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation, or both. Reproduction is authorised provided the source is acknowledged.

Table of Contents

Executive Summary.....	11
1 Introduction.....	12
1.1 Mapping NextGEM Outputs	12
1.2 Deliverable overview and report structure	12
1.3 Updates from previous Deliverable D6.2 “Network provisioning and links with EU health data space – First report”	13
2 Background and Motivation	14
2.1 Requirements for sharing of FAIR health data	14
2.1.1 Requirements for data collection, processing, and storage.....	14
2.1.2 Requirements for data sharing and distribution	14
2.1.3 Requirements for trust, interoperability, sovereignty, and scalability of network management	14
2.1.4 Requirements for access control	14
2.1.5 Requirements for FAIR data	15
2.1.6 Requirements for legal/regulatory issues	15
2.2 Data sharing in Europe	15
2.2.1 Existing data sharing architectures.....	15
2.2.2 Existing health data lakes and databases	16
2.2.3 Existing protocols and mechanisms for data sharing	17
2.2.4 Existing knowledge-sharing platforms	18
2.2.5 European projects on health data sharing.....	18
2.2.6 Access control in European and national initiatives on health	19
3 Data space ecosystem	20
3.1 Data space design principles	20
3.2 Existing architectures and the International Data spaces architectural model	21
3.3 Existing data space ecosystems.....	22
3.3.1 Health data space.....	23
3.3.2 Cognitive Ports data space.....	25
3.3.3 Automotive data space.....	25
3.3.4 Mobility data space	25
3.3.5 Energy data space	26
3.3.6 Manufacture data space.....	26
3.3.7 Earth observation data space.....	26
3.3.8 Agriculture data space	27
3.3.9 Finance data space.....	27
3.4 Data space mechanism and connectors	27
3.4.1 Data space protocol.....	27
3.4.2 Data space components	28
3.4.3 Data space connectors	28

3.5	Initial deployment of data space solutions.....	31
3.5.1	Single-connectors MVD	32
3.5.2	IDSA MVD.....	33
3.5.3	Sovity MVD	34
3.5.4	Eclipse MVD	36
3.6	Comparison of connector deployments.....	40
4	NextGEM network provisioning platform	41
4.1	Integration of data space principles in the NextGEM architecture.....	41
4.2	NextGEM data space topology	41
4.2.1	Data space premises	41
4.2.2	Data space connector.....	41
4.2.3	Data space data storage.....	42
4.2.4	Data space network	42
4.3	Data space orchestration	42
4.3.1	Controller	42
4.3.2	Data space orchestration.....	43
4.4	Interoperability between data spaces and services	44
5	NextGEM security and privacy	45
5.1	Data access controls and trustworthiness.....	45
5.1.1	FAIR.....	45
5.1.2	Policies	45
5.1.3	Access control.....	45
5.2	Security concerns and countermeasures in NextGEM platform.....	46
5.2.1	Component security	47
5.2.2	Network security.....	47
5.2.3	SPA Tool integration with NIKH components	47
5.2.4	SPA Tool vulnerability assessments and monitoring	48
6	NextGEM Services	51
6.1	Towards NextGEM data space implementation.....	51
6.2	Data-as-a-Service (DaaS).....	52
6.3	Connector-as-a-Service (CaaS).....	53
6.4	Premises-as-a-Service (PaaS).....	53
6.5	Dataspace-as-a-Service (DSaaS).....	54
7	Development of NIKH data space platform	55
7.1	Deployment overview and service topology.....	55
7.2	User interface	56
7.2.1	Assets	58
7.2.2	Contracts	61
7.2.3	Policies	62

7.2.4	Premises.....	63
7.2.5	Users	63
8	Evaluation of NIKH Data space platform in realistic scenarios	64
8.1	NextGEM Realistic Scenarios.....	64
8.1.1	Usage Scenario 1 – NextGEM Members	64
8.1.2	Usage Scenario 2 – NextGEM Local premises	64
8.1.3	Usage Scenario 3 – NextGEM Limited Local Premises	64
8.1.4	Usage Scenario 4 – External Data space ecosystem.....	65
8.2	Discussion	65
9	Conclusion.....	66
10	References.....	67

List of Figures

Figure 1: Core IDS components and their interactions	20
Figure 2: Interaction of the IDS technical components	22
Figure 3: European Health Data space initiative	23
Figure 4: The Data space protocol as a technical innovation and strategic enabler	27
Figure 5: An overview of data space connectors (IDSA)	29
Figure 6: IDS Connector data model	30
Figure 7: EDC Core Module System and its architecture	31
Figure 8: Sequence diagram of the Provider Push paradigm	32
Figure 9: Sovity MVD UI Dashboard	34
Figure 10: Time-Period-Restricted Policy Creation in the Sovity Connector UI	35
Figure 11: Contract Negotiation in the Sovity Connector UI	35
Figure 12: Transfer History list in the Sovity Connector UI	36
Figure 13: Eclipse Minimum Viable Data space deployment	37
Figure 14: Flow sequence diagram of distributed authorization between data space participants in MVD	37
Figure 15: Catalog browsing lists all the available assets within the data space	38
Figure 16: Blob storage and its contents	38
Figure 17: A contract definition associated with an asset	39
Figure 18: Transfer process	39
Figure 19: Consumer's storage account	40
Figure 20: Connector information is acquired dynamically	43
Figure 21: Asset transfer sequence	43
Figure 22: Shipping NIKH logs to the SPA tool	48
Figure 23: Example list of NVD assessment results	49
Figure 24: Example NVD finding	49
Figure 25: Availability Assessment overview	49
Figure 26: Availability Assessment Result & Rule	50
Figure 27: Premises – Connector – Data space, the three key framework components of the NIKH platform	51
Figure 28: The relation between all 'as-a-Service' scenarios	52
Figure 29: Deployment of NIKH data space ecosystem	55
Figure 30: NIKH's Front page	56
Figure 31: NIKH's login prompt	57
Figure 32: NIKH's Tools page	57
Figure 33: NIKH's Data space ecosystem page	58
Figure 34: NIKH's Assets view	59
Figure 35: NIKH's Asset details	59
Figure 36: NIKH's Create asset form	60
Figure 37: NIKH's Assets catalog view	61
Figure 38: NIKH's Contracts view	61

Figure 39: NIKH's Create contract form62

Figure 40: NIKH's Policies view.....62

Figure 41: NIKH's Premises view63

Figure 42: NIKH's Users view63

Figure 43: NIKH Platform data space services in NextGEM realistic usage scenarios64

List of Tables

Table 1: Adherence to NextGEM's GA Tasks and Deliverables Descriptions	12
Table 2: Comparison of connectors deployment	40
Table 3: User roles and permissions	46
Table 4: Data-as-a-Service specification	52
Table 5: Connector-as-a-Service specification	53
Table 6: Premises-as-a-Service specification	54
Table 7: Data space-as-a-Service specification	54
Table 8: Validation of Data space services in realistic scenarios	65

Glossary of terms and abbreviations used

Abbreviation / Term	Description
API	Application Program Interface
CI/CD	Continuous Integration/Continuous Development
CLUE-H	European Cluster EMF and Health
CVE	Common Vulnerabilities and Exposures
DMP	Data Management Plan
DOI	Digital Object Identifier
DPIAs	Data Protection Impact Assessments
DPO	Data Protection Officer
EMF	Electromagnetic Field
ERMES	Electric Regularized Maxwell Equations with Singularities
FAIR (principles)	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
GUI	Graphical User Interface
IA	Information Architecture
ICNIRP	International Commission on Non-Ionizing Radiation Protection
IDS	International Data space
IDSA	International Data Space Association
IOT	Internet of Things
JSON	JavaScript Object Notation
NextGEM	Next Generation Integrated Sensing and Analytical System for Monitoring and Assessing Radiofrequency Electromagnetic Field Exposure and Health
NIKH	NextGEM Innovation & Knowledge Hub
NVD	National Vulnerability Database
OpenAIRE	Open Access Infrastructure for Research in Europe
RA	Risk Assessment
REST	Representation State Transfer

RF	Radio Frequency
SPA	Security and Privacy Assurance
SLA	Service-Level Agreement
SOP	Standard Operating Procedures
TBD	To Be Defined
UC	Use Case
US	Usage Scenario(s)
UX	User Experience
VM	Virtual Machine
WP	Work Package

Executive Summary

Deliverable D6.7 reports on the activities of Task 6.2 “Network provision and links with EU health data space” which is part of WP6 on the “Development of NextGEM Innovation and Knowledge Hub”.

The primary goal of this deliverable is to describe the network provision mechanisms between the distributed data locations that are to be used throughout the duration of Task 6.2. This task provides a robust solution and access control on the NextGEM platform, with a particular focus on establishing and ensuring a secure environment with respect to the legal/regulatory issues as implemented in Task 6.3, aiming to address the challenging objective of creating a hub for Electromagnetic Fields (EMF) and Health.

The main contribution of this deliverable is the provision of secure network sharing mechanisms within NextGEM by describing the requirements, intended use, main functionalities, and the core sub-components that it is comprised of and deploying the data space mechanisms towards the development of the NextGEM data space ecosystem.

1 Introduction

The main scope of Deliverable D6.7 “Network provisioning and links with EU health data space – Final report” is to ensure reliable and scalable information sharing within the NextGEM for secure data between distributed storage locations and the health data space. The report overviews the requirements for health data sharing and existing initiatives on the European level. In addition, a comprehensive analysis of the data space principles and solutions is also included. Accordingly, NextGEM aims to create its own NextGEM data space ecosystem for secure and reliable network provisioning between distributed storage locations. Thus, the focus is on implementing the NextGEM health data space by applying robust access control mechanisms in compliance with European legal and regulatory standards and developing techniques for secure, privacy-conscious data access and sharing.

1.1 Mapping NextGEM Outputs

The purpose of this section is to map NextGEM’s Grant Agreement (GA) commitments, both within the formal Task description and Deliverable, against the project’s respective outputs and work performed.

Table 1: Adherence to NextGEM’s GA Tasks and Deliverables Descriptions

TASKS	
Task Number & Title	Respective extract from formal Task Description
Task 6.2 - Network provision and links with EU health data space.	This task offers network provision guarantees between the distributed data location as stored in the NextGEM platform including information sharing process in a scalable and reliable manner. More specifically, cloud environments are of key importance to preserve user’s data collection, analysis, processing, and storage in the compute continuum. In close cooperation with other Tasks of this WP, this task provides a robust solution to access control in NextGEM platform. Special attention is given to establish the basics for guarantee a secure environment with respect to the legal/regulatory issues currently existing in Europe as implemented in T6.3. In addition, this task researches and works on techniques and technologies to control access, with the aim of sharing the exported data NextGEM in secure and privacy presented manner. The outcome provides secure interfaces for NextGEM platform with relevant data spaces initiatives as promoted by GAIA-X and EU Health Data Spaces.
DELIVERABLE	
Deliverable: D6.7: Network provision and links with EU health data space – Final Report (M32) This deliverable provides the final report of network provision between the NextGEM platform and associated distributed partners location for reliable data sharing and exchange of data to the EU data space.	

1.2 Deliverable overview and report structure

Based on the objectives and work carried out under Task 6.2, the document starts with the Executive Summary followed by the introduction of the document in Section 1.

Section 2 provides a comprehensive overview of the background, highlighting the requirements that need to be fulfilled to meet the needs described in the usage scenarios, as well as the existing initiatives and data-sharing architectures.

Section 3 presents the data space ecosystem, describing the data space principles, and highlights the existing data space solutions.

Section 4 introduces the proposed topology for supporting network provisioning between distributed storage locations.

Section 5 outlines the security and privacy measures, including the SPA tool, implemented to assess and monitor the NIKH platform.

Section 6 presents an overview of the services provided by the platform.

Section 7 covers the containerized NIKH ecosystem for secure data exchange and details its user interface functions.

Section 8 evaluates the NIKH Data Space platform's performance in NextGEM scenarios.

Finally, Section 9 concludes the deliverable with a summary of outcomes and future actions.

1.3 Updates from previous Deliverable D6.2 “Network provisioning and links with EU health data space – First report”

As this deliverable continues from Deliverable D6.2 “Network provisioning and links with EU health data space – First report”, it builds upon the previous results and work undertaken, with the following additions:

- Sections 2 and 3 were updated to include additional European projects in health and further description of the European health data space.
- The latest table depicting an overview of data space connectors as per IDSA was added to Section 3.4.3.
- The user roles and permissions have been updated in Section 5.1.3 so that they include the roles “Administrator”, “Full Member”, and “Team Member”.
- Section 5.2.4 introduces enhanced SPA tool capabilities, building on D6.2's security framework. It details NVD-based vulnerability assessments (added Figures 23 and 24), availability monitoring of VMs hosting Data space services and external databases, and a planned penetration test for developer-shared remediation, with further insights in D6.3 and D6.8.
- Section 6 is updated with regards to the data space implemented scenarios and functions of use.
- The new Section 7 outlines the containerized, secure data space ecosystem of the NIKH platform and explains the robust access control and governance mechanisms to enable controlled, sovereign information exchange. It also provides a step-by-step description of the different functions offered via NIKH's user interface.
- The new Section 8 assesses the NIKH Data Space platform's real-world performance through diverse NextGEM scenarios, emphasizing secure and scalable data sharing. It evaluates different deployment models and their operational challenges in ensuring effective data governance and access within a health data ecosystem.

2 Background and Motivation

2.1 Requirements for sharing of FAIR health data

The Findable, Accessible, Interoperable, and Reusable (FAIR) sharing of health data in the European Union is governed by several requirements that are aimed at ensuring data privacy, security, and interoperability while facilitating data sharing for research and healthcare purposes.

2.1.1 Requirements for data collection, processing, and storage

The collection, processing, and storage of health data must comply with the General Data Protection Regulation (GDPR) ¹. This required transparency and lawful data collection practices, including obtaining an informed consent form (ICF) from volunteers and ensuring data minimization and purpose limitation. Moreover, high data quality is essential to maintaining the accuracy, completeness, and reliability of health data. This requires of inter alia the implementation of protocols using high-quality methodologies, such as clear inclusion/exclusion criteria, controlled exposure conditions relevant to the needs of the experiment, including dosimetry, blinded experiments and appropriate controls.

Moreover, data processing and storage must follow security measures to mitigate the risk of unauthorised access or disclosure. To this end, personal data will be anonymized, and only the coded datasets will be made available. These datasets will not contain any details enabling individuals to be identified, in compliance with current regulations on the protection of privacy (GDPR). All published results will lack any information that could identify individuals' identities. The list of participants and their respective codes will be kept on separate secure servers, access to which will be restricted to the identified study managers, as defined in documents approved by the local ethics committee. The process is supervised by the NextGEM's ethics board.

2.1.2 Requirements for data sharing and distribution

Data sharing and distribution need to follow requirements designed to facilitate responsible data sharing while protecting individuals' rights and promoting trust in the use of data. As defined in Section 2.1.1, only pseudonymized personal health data will be shared, while the respective codes will be kept closed and stored on a separate secure server. Access to servers are managed under the responsibility of the partner institutions involved in studies entailing health data, in agreement with their Data Protection Officer (DPO).

2.1.3 Requirements for trust, interoperability, sovereignty, and scalability of network management

The sharing of pseudonymized health data involves defining clear guidelines, protocols, and certifications for data sharing and network management, ensuring transparency, accountability, and reliability in data-handling practices. Moreover, it is needed to promote interoperability by defining consistent data formats, structures, and communication protocols. Building a scalable infrastructure that can accommodate the growing volume and complexity of data is essential. The NIKH will implement mechanisms to ensure the following key principles:

- **Trust:** data providers and consumers need to be confident in the network's security and reliability, implementing mechanisms for authentication, authorization, and access control.
- **Interoperability:** enables data space networks to involve diverse participants with a variety of systems and protocols, ensuring data exchange through seamless communication.
- **Sovereignty:** enables data owners to control their data, allowing them to self-determine how their data is collected, stored, shared, and used by others.
- **Scalability:** supports the handling of massive amounts of data efficiently.

2.1.4 Requirements for access control

Access control is essential for ensuring that only authorized individuals or entities can access sensitive health data. Several requirements govern access control to protect data privacy, security, and confidentiality while enabling legitimate users to retrieve and utilize necessary information, such as the implementation of a robust user authentication mechanism, which ensures that only authenticated users can access the network and its resources.

Moreover, access control policies are established to define rules and criteria for granting or denying access to data and resources. Access control policies specify who can access what data, under what conditions, and for what

¹ <https://gdpr-info.eu/>

purposes, ensuring that access rights are managed consistently and transparently across the network. Enabling access revocation mechanisms will allow administrators to revoke access privileges promptly when users no longer require access or when security incidents occur. The processes are managed under the responsibility of the partner institution involved in studies entailing health data, in agreement with their DPO.

2.1.5 Requirements for FAIR data

Seamless collaboration and information sharing, especially among European Health stakeholders, have emerged as a crucial need within the scientific community, fostering knowledge discovery and innovation [1]. However, storing and sharing heterogeneous data is a challenging task and requires the development of complex solutions. To this end, FAIR principles are able to tackle these challenges, enhancing the ability of the stakeholders to find and use data from multiple sources [2]. To achieve this, comprehensive metadata standards and documentation must accompany datasets and provide information on the data's origin, structure, and usage guidelines. Several requirements need to be implemented to promote the FAIR data:

- **Findability** of the datasets and other project outputs is crucial to enable stakeholders to discover and access data efficiently for analysis, research, and decision-making purposes. To enhance findability, providing detailed metadata that includes information such as the data's title, creator, description, keywords, persistent identifiers licensing, and contacts is crucial. Adhering to standardized metadata formats and vocabularies and using common metadata schemas are furthermore good practices to improve findability. The use of a searchable metadata catalog plays also a role in improving the data's findability.
- **Accessibility** means ensuring that data is readily available and accessible to users or automated systems. Access policies need to be implemented to facilitate data accessibility by making datasets as open as possible and as closed as necessary, as set by the European Commission. Using open, non-proprietary data formats and standards promotes data accessibility by ensuring compatibility and interoperability across different systems and disciplines.
- **Interoperability** refers to the ability of data to be exchanged, integrated, and used seamlessly across different systems, platforms, and disciplines. Achieving interoperability involves using standardized data formats. These standards ensure that data can be understood and interpreted consistently across different systems and disciplines.
- **Reusability**: Standardized formats facilitate data exchange and reuse, enabling users to access and interpret data without specialized software or tools.

2.1.6 Requirements for legal/regulatory issues

These requirements ensure that health data is used in compliance with various laws, regulations, and ethical guidelines to guarantee the legal and ethical use of health data. It means ensuring compliance with the GDPR to protect individuals' privacy rights and regulating the processing of personal data, including health data. This involves obtaining ICFs from volunteers for data processing, implementing data protection measures, and adhering to principles such as data minimization and purpose limitation. Compliance is ensured through the supervision of local ethics committees, in agreement with the DPOs of the institute in charge of managing this sensitive data, and also under the supervision of NextGEM's ethical board.

2.2 Data sharing in Europe

2.2.1 Existing data sharing architectures

Various data lake architectures have been proposed to facilitate the ingesting, storing, and sharing of heterogeneous data, enabling users to explore and analyze the available data [3]. The authors in [4] introduce the Azure Data Lake Store (ADLS), a fully managed, elastic, scalable, and secure file system that encapsulates the Hadoop distributed file system (HDFS) and Cosmos semantics. It is specifically designed and optimized for a wide range of big data analytics that depend on a high degree of parallel reads and writes, providing security and data-sharing features. In [5], the authors introduce a platform designed to address the need for data sharing of heterogeneous scientific data related to environmental and agricultural research. The platform employs an on-premises solution utilizing the data lake concept to provide cloud services with both institutional authentication and open data access. The article in [6] presents an online data lake that allows users to collect, store, analyze, and share heterogeneous cultural heritage data. It proposes a zero-administrator, zero-cost, integrated framework that enables researchers and other stakeholders to (i) deploy data acquisition services (ii) design and manage versatile customizable data stores, (iii) share whole datasets or horizontal/vertical data with other stakeholders, (iv) search, filter and analyse data via an expressive yet simple-to-use graphical query engine and visualization tools, and (v) perform user management and

access control operations on the stored data. The authors in [7] propose the HEALER, a data lake architecture that effectively performs data ingestion, storage, and access with the aim of providing a centralized repository for efficient storage of heterogeneous health data. The proposed architecture supports the ingestions of various data, performs waveforms processing to make them more interpretable to researchers, grants access to the saved data and allows the analysis of natural language reports. However, traditional data sharing platforms, which act as authoritative third parties for transactions, should be trusted by all participants in the sharing process.

In a centralized architecture, access control is usually the responsibility of the service provider. To overcome this limitation, some studies have proposed schemes that replace these centralized architectures with blockchain-based solutions, providing trustworthiness, decentralization, immutability, and auditability. Blockchain technology is a decentralized digital ledger that records transactions across a network of computers in a way that ensures the data is secure, transparent, and immutable. Each transaction is grouped into a "block" and linked to the previous block, forming a "chain." This structure prevents alteration of any previous transactions without altering all subsequent blocks, which requires consensus across the network. Therefore, the authors of [8] propose a blockchain-based solution that can provide a more intensive mechanism for data access control, enabling data owners to specify the access policies, and the schemes to achieve fine-grained access control over data. In [9], authors present the MeDShare, a blockchain-based solution that leverages smart contracts, access control and active monitoring mechanisms to provide a secure and transparent framework for sharing sensitive data among participants, aiming to mitigate privacy risks and unauthorized access in a trustless environment. In [10] a Hyperledger-based sharing architecture was proposed that facilitates (i) private and auditable healthcare data sharing and (ii) healthcare data access and permission handling by leveraging inherent properties of the blockchain technology (e.g., immutability, auditability, and accountability) combined with the usage of smart contracts, a transaction-aware state machine mechanism that associates a medical record with viewing permissions and data retrieval instructions. In the medChain project [11], authors proposed a Blockchain-based architecture for a healthcare data-sharing schema that ensures privacy using encryption with private key and hashing of sensing data before being stored on the medChain Blockchain. This solution provides a data-sharing schema that achieves security, integrity, auditability and privacy-preservations goals. Despite the promising features of blockchain and smart contracts-based solutions in the context of providing security for sensitive data, these technologies still face some challenges limiting their extensive usage for data sharing [12]. Blockchain networks can struggle with scalability issues from the perspective of throughput, storage and networking. Actually, as the number of participants increases, the distributed ledger increases drastically make it unusable for complex and large scale implementations.

2.2.2 Existing health data lakes and databases

2.2.2.1 International health databases

International health databases represent collaborative initiatives aimed at aggregating health data at an international level to facilitate cross-border research and comparative analyses. A prominent example is the World Health Organization (WHO) Data Hub ², which is a digital platform for global health data, allowing for data collection, storage, sharing and analysis. In the context of health data regarding the potential health impacts of EMF, another example is the WHO International EMF Project ³, which provides a platform for the collection, analysis, and dissemination of data on the health effects of electromagnetic fields through The Global Health Observatory ⁴. Through its collaboration with Member States and expert groups, the project provides a repository of scientific literature, exposure assessments, and health risk evaluations related to EMF, supporting evidence-based decision-making and risk communication at a global level.

2.2.2.2 European health databases

European health databases comprise initiatives aimed at harmonizing health data across EU Member States to support regional collaboration and policy development, with some initiatives addressing EMF exposure. They include patient records, clinical trial data, epidemiological studies, and pharmaceutical research. Examples include the European Health Data Space (EHDS), Eurostat health statistics, and the European Medicines Agency (EMA) databases. These databases aim to improve public health, facilitate medical research, and support policy-making while ensuring data privacy and security under regulations like GDPR. The European Health for All Database

² <https://data.who.int/>

³ <https://www.who.int/initiatives/the-international-emf-project>

⁴ <https://www.who.int/data/gho>

(HFA-DB) ⁵, available through the WHO European Health Information Gateway, provides comprehensive health statistics for the WHO European Region. It includes data on demographics, health status, risk factors, healthcare resources, and expenditures from over 50 countries. The database helps policymakers, researchers, and public health professionals analyze trends and make evidence-based decisions. It supports monitoring progress towards health-related Sustainable Development Goals (SDGs) and other international commitments. The data is regularly updated and follows WHO's standards for accuracy and comparability across countries.

2.2.2.3 National health databases

National health databases comprise repositories of health data maintained by individual countries to support domestic healthcare delivery, policy formulation, and research. One such example is IDIKA ⁶, a Greek platform that provides innovative eHealth solutions, focusing on the development and deployment of national health IT systems. IDIKA supports the secure sharing and management of health data across various stakeholders in the healthcare system and implements strict data privacy policies to ensure the protection and confidentiality of data. The Federal Health Monitoring Information System (IS-GBE)⁷, managed by the Federal Statistical Office (Destatis), is particularly prominent. This online platform provides public access to regularly updated health data from over 100 different sources, covering topics such as hospital statistics and causes of death. Its comprehensive nature and accessibility make it a cornerstone for health-related data in Germany. In Sweden, the National Patient Register (NPR) ⁸ provides detailed records of inpatient and outpatient care, aiding medical research. Italy's National Health Information System ⁹ integrates hospital and pharmaceutical data to improve healthcare planning. Meanwhile, Spain's Sistema de Información Sanitaria del SNS (SIS-SNS) ¹⁰ compiles nationwide health records to enhance disease surveillance and patient care. These databases contribute significantly to European health initiatives and cross-border collaborations.

2.2.3 Existing protocols and mechanisms for data sharing

Health data-sharing protocols and mechanisms facilitate the interoperability, standardization, and secure exchange of health information among systems and stakeholders. Leveraging such standards and mechanisms enable health-related organizations to facilitate seamless data exchange and support data-driven decision-making. Notable protocols and mechanisms for data sharing are given below.

Fast Healthcare Interoperability Resources (FHIR) ¹¹ developed by the HL7 International organization is a leading standard for healthcare. FHIR leverages tools such as RESTful (Representational State Transfer) APIs to enable integration and interoperability across healthcare systems. Its standardized data model and resource-centric approach enable efficient and scalable data exchange for a diverse range of use cases, such as patient care coordination and clinical research.

Health Level Seven (HL7) ¹² standards encompass interoperability specifications governing the exchange, integration, and management of electronic health information. HL7 standards, including HL7 Version 2 (HL7 v2) and HL7 FHIR, serve as foundational frameworks for healthcare data exchange, enabling communication between disparate systems and stakeholders and allowing effective data sharing across healthcare ecosystems.

Substitutable Medical Applications, Reusable Technologies (SMART) ¹³ extends the FHIR standard to enable the development of interoperable healthcare applications and plug-and-play solutions. By leveraging OAuth 2.0 for authentication and authorization, SMART on FHIR enables secure access to electronic health record (EHR) data and integration with third-party applications. Through its modular and extensible architecture, SMART on FHIR allows developers to build healthcare applications that integrate with existing clinical workflows.

⁵ <https://gateway.euro.who.int/en/datasets/european-health-for-all-database/>

⁶ <https://www.idika.gr/>

⁷ <https://www.gbe-bund.de/>

⁸ <https://www.socialstyrelsen.se>

⁹ <https://www.salute.gov.it/portale/lea/dettaglioContenutiLea.jsp?area=Lea&id=5073&lingua=italiano&menu=vuoto>

¹⁰ <https://www.sanidad.gob.es/estadEstudios/estadisticas/sisInfSanSNS/tablasEstadisticas/home.htm>

¹¹ https://ecqi.healthit.gov/fhir?qt-tabs_fhir=about

¹² <https://www.hl7.org/>

¹³ <https://docs.smarthealthit.org/>

2.2.4 Existing knowledge-sharing platforms

NextGEM recognizes the importance of leveraging existing data-sharing platforms to promote collaboration, facilitate access to relevant datasets, and share its findings with the broader scientific community and directly involved stakeholder groups, such as regulatory bodies. The NIKH Platform will enable bilateral data access with such data-sharing platforms through APIs. The following platforms have been identified as key resources.

Zenodo¹⁴ was established in 2013 through a collaboration between OpenAIRE and CERN and serves as an open-access repository dedicated to facilitating the sharing and archiving of research outputs across diverse scientific disciplines. As an integral component of the European Open Science Cloud (EOSC), Zenodo embodies the principles of openness, reproducibility, and long-term data preservation, thereby aligning closely with the ethos of the NextGEM project.

Dataverse¹⁵ is developed by the Institute for Quantitative Social Science at Harvard University, and emerged as a pioneering platform for the dissemination, preservation, citation, and analysis of research data. Since its inception in 2006, Dataverse has evolved into a robust data management solution, providing researchers with a standardized framework for depositing, discovering, and collaborating on datasets while adhering to best practices in data sharing and stewardship.

Yoda¹⁶ was initiated by the Yale University Open Data Access (YODA) Project in 2011, and represents a seminal platform for the sharing of individual participant data derived from clinical trials. Grounded in principles of transparency, accountability, and data sharing, Yoda enables researchers to access anonymized participant-level data, fostering secondary analyses and advancing biomedical research endeavors.

EMF-Portal¹⁷ was established through collaborative efforts between RWTH Aachen University and the University Medical Centre Freiburg in 2004 and serves as a comprehensive repository of scientific literature pertaining to the biological effects of electromagnetic fields (EMFs). Leveraging rigorous curation practices and peer-reviewed content, EMF-Portal facilitates evidence-based decision-making and risk assessment within the realm of EMF exposure research.

The **Web of Science platform**¹⁸ was introduced by the Institute for Scientific Information (ISI) in 1964, and stands as a preeminent multidisciplinary citation database encompassing scholarly literature across various academic domains. Renowned for its extensive coverage of peer-reviewed journals and conference proceedings, Web of Science remains a cornerstone resource for researchers seeking to access, evaluate, and contextualize scientific literature within their respective fields of inquiry.

PubMed¹⁹ was established by the National Centre for Biotechnology Information (NCBI) at the National Library of Medicine (NLM) in 1996, and represents a seminal platform for accessing biomedical literature. As a freely accessible search engine indexing a vast repository of peer-reviewed research articles, PubMed facilitates evidence-based decision-making and fosters advancements in biomedical research through its comprehensive bibliography.

2.2.5 European projects on health data sharing

EU-funded projects have been pivotal in facilitating health data sharing across the continent, without relying exclusively on the Data space concept. These projects aim to enhance data interoperability, security, and accessibility, while ensuring compliance with data protection regulations. The following projects are a non-exhaustive list of projects regarding health data sharing.

The **European eHealth Interoperability Conformity Assessment Scheme for Europe (EURO-CAS)**²⁰ is a Horizon 2020 project that developed a sustainable Conformity Assessment Scheme to ensure the interoperability

¹⁴ <https://zenodo.org/>

¹⁵ <https://www.iq.harvard.edu/computer-labs/lab-resources/research/dataverse>

¹⁶ <https://yoda.yale.edu/>

¹⁷ <https://www.emf-portal.org/en>

¹⁸ <https://clarivate.com/products/scientific-and-academic-research/research-discovery-and-workflow-solutions/webofscience-platform/>

¹⁹ <https://pubmed.ncbi.nlm.nih.gov/>

²⁰ https://hope.be/EU_Projects/euro-cas/

of eHealth systems across Europe. The project focused on creating a standardized framework for assessing the interoperability of health IT systems, ensuring that they can effectively communicate and share data.

MyHealthMyData (MHMD) ²¹ project is a Horizon 2020 that leverages blockchain technology to create a secure and transparent data-sharing framework for personal health data. MHMD aims to enhance the control individuals have over their data while facilitating secure and efficient data exchange among healthcare providers, researchers, and patients. This approach reduces the reliance on data repositories, enhancing data security and sovereignty.

Participatory Urban Living for Sustainable Environments (PULSE) ²² is a Horizon 2020 project that focuses on using big data and IoT technologies to improve public health outcomes in urban areas. By collecting and analysing health and environmental data from various sources, PULSE supports evidence-based decision-making and promotes healthier urban living. The project emphasizes data interoperability and real-time data sharing to address public health challenges effectively.

FAIR4Health ²³ is an EU-funded project aims to implement the FAIR data principles in health research data management. FAIR4Health supports the sharing and reuse of health research data by developing tools and guidelines that facilitate data interoperability and compliance with FAIR principles. The project enhances the accessibility and usability of health research data across Europe.

SEAWave ²⁴ is an EU-funded project, that contributes to the scientific basis for 5G health risk assessment and provides the means for effective health risk communication. SEAWave focuses on identifying knowledge gaps in RF-EMF exposure, including differences between 5G and previous generations, exposure in workplaces, new beamforming technologies, dosimetry of the human skin, health risk studies, assessment of end-user devices, and public perception of 5G exposure. The project aims to provide tools for reliable exposure evaluation and contribute to scientific knowledge on potential health risks.

ETAIN ²⁵ is an EU-funded project that aims to assess the impact of RF-EMF from both a human and planetary health perspective while exploring options for exposure reduction. The project develops an EMF Monitor app and portal to visualize RF-EMF exposure levels across Europe. Key objectives include implementing an integrated dose model, providing recommendations for exposure reduction, assessing RF-EMF effects on biodiversity (particularly insects), studying long-term health impacts on skin and eyes, evaluating planetary health implications, and engaging stakeholders through citizen-science approaches.

GOLIAT ²⁶ is an EU-funded project that aims to respond to some of the questions raised by the new wireless technologies, with a special focus on 5G. GOLIAT aims to monitor RF-EMF exposure, particularly from 5G, provide novel insights into its potential causal health effects, and understand how exposures and risks are perceived and best communicated using citizen engagement. Results will be published in open-access peer-reviewed journals through which potential users will become aware of the data generated by GOLIAT. After publication of the results, data will be made available to external researchers to maximize their impact by encouraging secondary analyses to address other research questions.

2.2.6 Access control in European and national initiatives on health

Access control mechanisms play a critical role in safeguarding the privacy, security, and integrity of health data within European and national healthcare initiatives. This chapter examines the access control strategies employed in patient health records, healthcare provider systems, and other healthcare-related platforms at both the European and national levels. The cornerstone of personal data protection for individuals in the EU lies in the GDPR. GDPR has been enforced across the EU since 2018 and establishes regulations governing the processing and protection of personal data, including biometric and health data. Under the GDPR, healthcare organizations must implement robust access control measures to ensure that only authorized individuals, such as healthcare professionals involved in patient care, have access to sensitive health data. Access controls must be aligned with the principles of data minimization, purpose limitation, and confidentiality to safeguard patient privacy and confidentiality.

²¹ <https://www.myhealthmydata.eu/>

²² <https://www.project-pulse.eu/>

²³ <https://www.fair4health.eu/>

²⁴ <https://seawave-project.eu/>

²⁵ <https://www.etaingroup.eu/>

²⁶ <https://projectgoliat.eu>

3 Data space ecosystem

A data space ecosystem consists of Connectors that link data providers and data consumers with other components, such as identity management components (i.e., the Certificate Authority, Dynamic Attribute Provisioning Service, and Participant Information Service). These components manage identity information for participants in the data space ecosystem in order to avoid unauthorized access to data. The Metadata Broker acts as an intermediary that manages the metadata repository, providing information about the data sources. From a technical perspective, some components are optional, such as the Broker, whereas the Connectors, as well as the components related to the Identity management, are mandatory to create a data space and establish a secure and sovereign data exchange, as can be seen in Figure 1.

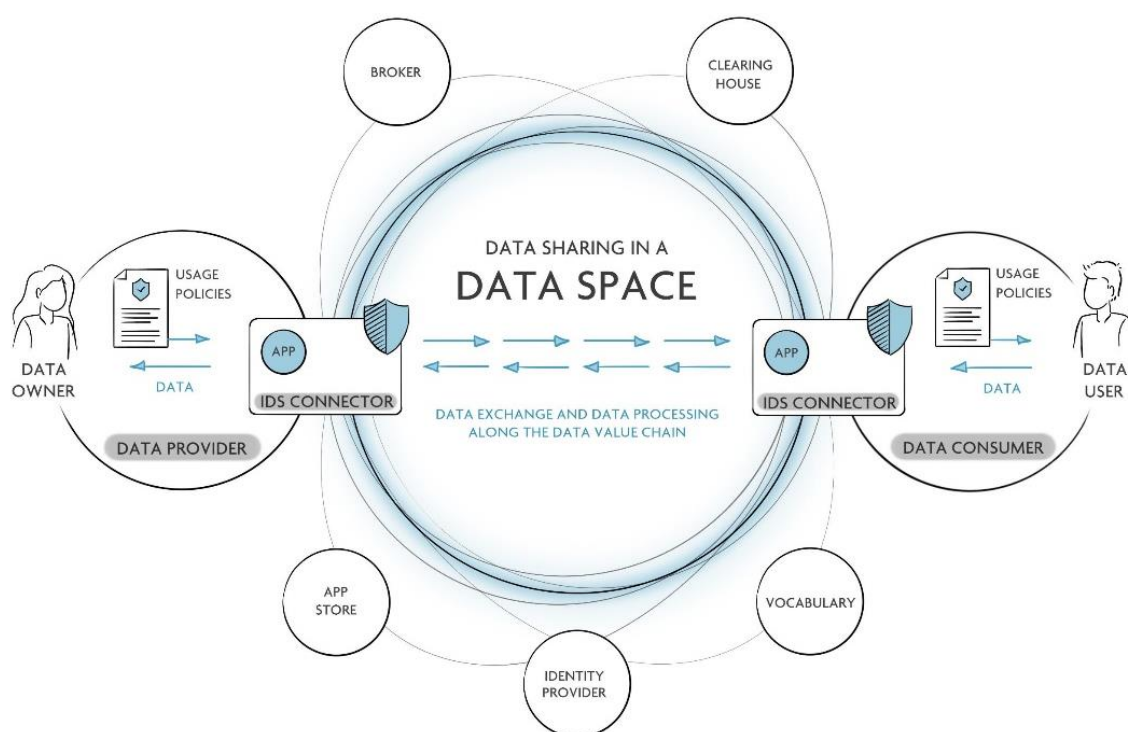


Figure 1: Core IDS components and their interactions ²⁷

3.1 Data space design principles

- **Trust.** A data space ecosystem tries to establish the policies and rules for reliable and trusted data sharing between different organizations. Data usage control is achieved through the appropriate mechanism implemented by Connectors, enabling data providers to keep control over the processing of their data by utilizing usage control policies.
- **Interoperability.** Designing a data space with semantic interoperability leads to more connected data and better utilized information across participants. In this direction, a data space ecosystem tries to facilitate interoperable data sharing between all the entities involved. Common standards related to participants and services ensure a level of compatibility (e.g., all entities understand the same machine-readable metadata). Another pillar of interoperability is the common trust framework. In this context, services can be certified by the same authority and share compliance services, using trust anchors for identification, enabling participants' identification in different data spaces using the same credentials. There are two main aspects of data space interoperability: (i) intra-data space interoperability, which ensures different participants within a single data space can exchange data effectively, even if they use different connectors, and (ii) inter-data space interoperability, which allows data spaces to connect and share data across them.
- **Sovereignty.** Regarding data, the concept of sovereignty deals with the need for a data owner to have full control over their own data when sharing these data with a third party. Consequently, in data-sharing ecosystems, data sovereignty is one of the core aspects to be preserved. Data usage control must provide

²⁷ <https://internationaldataspaces.org/why/data-spaces/>

mechanisms for respecting and protecting the data of all parties involved. Data providers must have access to monitoring and configuration tools that allow them to control what happens to their data.

- **Scalability and flexibility.** In real-world scenarios, in which the volume of the data is increased, the mechanisms designed to ensure data sovereignty have to meet some performance prerequisites. Usage policy evaluation mechanisms should be implemented in a reliable manner that will allow an efficient data exchange process. Moreover, the designed solution needs to be extremely flexible to conform to the variety of requirements presented in healthcare scenarios.

3.2 Existing architectures and the International Data spaces architectural model

The **International Data Space (IDS)** is a virtual environment built on existing standards, technologies, and governance models. It facilitates secure data exchange and data-based services among various stakeholders while ensuring data sovereignty. By proposing an architecture for secure data exchange and trusted data sharing, the International Data Spaces contributes to the design of enterprise architectures in commercial and industrial digitization scenarios. The IDS reference architecture model consists of the following layers.

- **Functional layer:** The Functional layer defines the functional requirements of a data space and the features that should be implemented.
- **Process layer:** The Process layer defines the interactions required between different components of the data space for achieving: (i) the onboarding process, in which access is granted to a data provider or data user, enabling access to data space resources, (ii) the exchanging data process, which includes searching capabilities for a data provider, invoking the actual data operation, and (iii) the publishing and using data, i.e., interacting with an IDS App Store or using IDS Data Apps.
- **Information layer:** The Information layer specifies the information model, an essential agreement between the participants that facilitates compatibility and interoperability. The primary purpose of this formal model is to enable (semi-)automated exchange of digital resources within a trusted ecosystem of distributed parties while preserving the data sovereignty of Data Owners.
- **System layer:** On the System Layer, the roles specified on the Business Layer are mapped onto a concrete data and service architecture to meet the requirements specified on the Functional Layer, resulting in what can be considered the technical core of the International Data Spaces. From the requirements identified on the Functional Layer, the major technical components result: the Broker, the Connector and the App Store. How these components interact with each other is illustrated in Figure 2.
- **Business layer:** The Business layer defines and categorizes the different roles that the participants of the data space ecosystem may have. Additionally, it specifies basic patterns and interactions between these roles, contributing to the development of innovative business models and data-driven services used within the framework of a data space.

GAIA-X²⁸ is a European initiative aimed at developing a federated and secure data infrastructure for Europe. It seeks to create a unified data ecosystem that ensures data sovereignty and promotes innovation across various industries. Figure 2 illustrates the interaction of technical components. The key aspects of GAIA-X are described as follows.

- **Data Sovereignty:** GAIA-X focuses on ensuring that data remains under the control of its owners. This means that businesses and individuals can manage, share, and process their data in a way that complies with European standards and regulations, such as the GDPR.
- **Interoperability and Portability:** The initiative aims to create a standardized framework that enables seamless data sharing and integration across different platforms and services. This includes interoperability between various cloud providers and the portability of data and services.
- **Federated Infrastructure:** GAIA-X promotes a federated approach where data remains distributed across different nodes. These nodes can be managed by different entities but are interconnected and adhere to common standards and policies.
- **Trust and Security:** GAIA-X prioritizes high levels of security and trust, ensuring that data is protected against unauthorized access and breaches. It establishes a set of rules and protocols that participants must follow to maintain the integrity and confidentiality of data.
- **Innovation and Competitiveness:** By providing a robust and secure data infrastructure, GAIA-X aims to foster innovation and increase the competitiveness of European companies. It supports various sectors,

²⁸ <https://gaia-x.eu/>

including industry, healthcare, finance, and public services, by enabling advanced data analytics and AI applications.

- **Collaborative Governance:** The initiative involves a wide range of stakeholders, including governments, businesses, research institutions, and non-profit organizations. This collaborative governance model ensures that the needs and interests of different parties are considered and balanced.

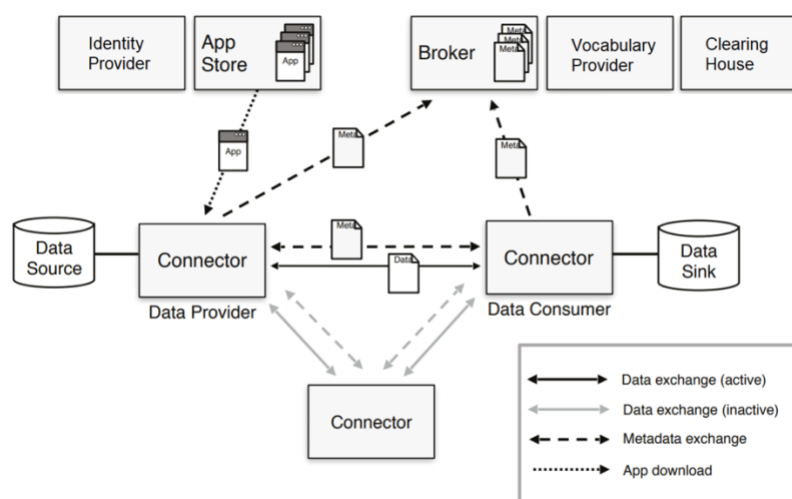


Figure 2: Interaction of the IDS technical components ²⁹

3.3 Existing data space ecosystems

The European strategy for data ³⁰ aims to create a single market for data, ensuring Europe's global competitiveness and data sovereignty. The concept of European data space fosters an ecosystem that enables the sovereign and self-determined exchange of data between different stakeholders, based on a common governance framework. In this context, Common European data spaces are specific data spaces that adhere to EU regulations and contribute to the objective of creating a single European market for the data. The key features of Common European data spaces include: (i) a federated data architecture that requires no central authority for data management or data governance tasks, enabling a secure and privacy-preserving framework for storing, accessing and sharing data, and (ii) a governance framework that aims to define rules in line with EU legislations taking into account the characteristics and idiosyncrasies of each specific sector. The Common European data spaces focus on the following sectors.

- A **Health Data Space** essential for health-related data, enabling secure sharing and access for research, personalized medicine, and healthcare improvements.
- An **Agriculture Data Space** essential for enhancing the sustainability and competitiveness of the agricultural sector, supporting the processing and analysis of relevant data.
- A **Manufacturing Data Space** essential for enhancing manufacturing processes, supply chains, and industrial innovation through secure data sharing.
- An **Energy Data Space** crucial for promoting efficient energy management, grid optimization, and renewable energy solutions.
- A **Mobility Data Space** essential for fostering transportation and mobility data, contributing to smart cities, logistics, and sustainable intelligent transport systems.
- A **Financial Data Space** essential for supporting financial services, risk assessment, and economic growth stimulating innovation, market transparency, sustainable finance, as well as access to finance for European businesses and a more integrated market.

²⁹ https://link.springer.com/chapter/10.1007/978-3-031-47444-6_3

³⁰ <https://digital-strategy.ec.europa.eu/en/policies/strategy-data>

- A **Public Administration Data Space** essential for enhancing public services, governance, and transparency through data collaboration.
- A **Skills Data Space** essential for skills development, education, and workforce planning.
- A **European Open Science Cloud (EOSC)** essential for providing the researchers with seamless access to scientific data.
- A **Green Deal Data Space** crucial for aligning with the European Green Deal objectives, emphasizing issues such as climate change, circular economy, pollution, biodiversity, and deforestation.

Additionally, data spaces related to media, cultural heritage, and smart communities have also emerged, contributing to the interconnected European data economy. The EU is also actively funding initiatives such as the **Data Spaces Support Centre** and **Smart Open-source Middleware (SIMPL)** to further advance these data spaces.

3.3.1 Health data space

A **Health Data Space** is a digital ecosystem designed to securely manage, store, and share health-related data across various entities like healthcare providers, researchers, and patients. It enables the seamless exchange of health information through standardized protocols while ensuring data security and patient privacy. This space facilitates better healthcare delivery, enhances medical research, and empowers patients to control their own health data. By promoting data interoperability, a Health Data Space supports more efficient, collaborative, and evidence-based healthcare practices.

3.3.1.1 The European health data space

The **European Health Data Space (EHDS)** ³¹, introduced by the European Commission in 2020, aims at fostering the sharing and reuse of health data across the European Union. EHDS is a health-specific ecosystem that fosters a genuine single market for health record systems, relevant medical devices, and high-risk AI systems (Figure 3). This ecosystem represents the first common EU data space in a specific area, emerging from the broader European strategy for data. The EHDS provides a health-specific data sharing framework that establishes common rules, standards, practices and infrastructures for the use of health data. In this direction, EHDS supports healthcare delivery, enabling secure data exchange, use and reuse of personal health data, and providing a framework for research and innovation. Thus, it contributes to evidence-based policymaking, fostering a consistent, trustworthy, and efficient setup for the use of health data in regulatory activities.

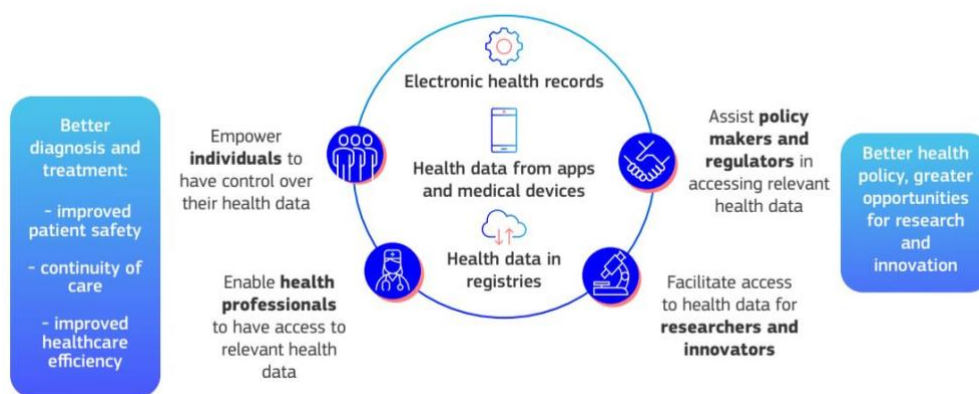


Figure 3: European Health Data space initiative ³²

Launched as part of the EU's digital strategy, the EHDS aims to create a unified and efficient data-sharing environment for healthcare and research, allowing for more effective healthcare delivery, innovation, and policy development. One of the central goals of the EHDS is to empower individuals to have more control over their own health data. By providing citizens with the ability to access and manage their health information digitally, the initiative enhances transparency and trust in healthcare systems. Citizens can grant or withdraw consent for their data to be used by healthcare providers, researchers, and policymakers, ensuring privacy and autonomy.

³¹ <https://www.european-health-data-space.com>

³² https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en

For healthcare providers, the EHDS offers a streamlined approach to exchanging patient information across borders. This is particularly beneficial in cross-border healthcare scenarios, such as when patients seek medical treatment in another EU country. It also enables healthcare professionals to access a patient's health records quickly, improving diagnosis, treatment, and care coordination.

Although the EHDS initiative is primarily focusing on the sharing and reuse of health data across the European Union, it is also engaged with compliance with data protection regulations. Access control mechanisms within the EHDS are designed to grant authorized users, such as healthcare providers, researchers, and public health authorities, appropriate access rights to health data based on their roles and responsibilities.

The EHDS also has significant implications for medical research. By facilitating the secure and anonymized sharing of health data across the EU, the EHDS creates new opportunities for collaborative research, particularly in areas like epidemiology, personalized medicine, and rare diseases. This data-driven approach can accelerate the development of innovative treatments and therapies, benefiting patients and healthcare systems across Europe.

Additionally, the EHDS supports the implementation of the European Health Data Governance Framework, which establishes clear rules for data access, use, and protection, ensuring that data is shared in a manner that respects privacy and ethical standards. By aligning the health data ecosystems of different EU countries, the EHDS aims to enhance cooperation, improve health outcomes, and ensure that the EU remains at the forefront of healthcare innovation and digital transformation.

In the long run, the EHDS will help to create a European Health Union by fostering closer integration between national health systems and creating a more resilient and adaptive healthcare infrastructure across Europe. Through better use of data, it can help address healthcare challenges such as aging populations, disease prevention, and the efficiency of healthcare delivery. Its primary goal is to enable better access to health data for research, policy-making, and healthcare delivery while ensuring privacy and data protection. The EHDS will allow individuals to have greater control over their health data and enable healthcare providers to share information across borders, improving the quality of care. By facilitating the use of health data for research, the initiative also aims to support innovation in the medical field and enhance public health strategies across Europe.

By promoting cross-border collaboration and facilitating data-driven healthcare innovations, EHDS seeks to enhance public health outcomes and facilitate the development of personalized medicine, thereby aligning synergistically with the overarching objectives of the NextGEM project to ensure the safety of EU citizens employing EMF-based telecommunication technologies.

3.3.1.2 *Other European health data space initiatives*

HEALTH-X dataLOFT³³ aims to implement this framework, offering a federated cloud infrastructure, which puts the individual at the centre of the access and control of future healthcare services. The provided use cases are summarized as follows:

- **Preventive care and treatment:** Citizens have the authority over their health data and benefit from improved preventive measures and better treatment options.
- **Research participation:** The platform enables citizens to actively engage in research projects.
- **Innovation hub:** Continual exploration of use cases ensures the platform's effectiveness and value for citizens and innovative business models.

HEALTH-X dataLOFT adheres to Gaia-X standards and aims to integrate health data from both the 'first health market' i.e., the area of "classic" healthcare, which is largely financed by statutory (GKV) or private health insurance (PKV), and 'second health markets', defined as all privately financed products and services.

Within the health context, the **HealthData@EU Pilot**³⁴ is another significant initiative aimed at overcoming existing barriers to the full utilization of digitalized health data within the European Union. The project is part of the broader EHDS initiative, fostering digital transformation and optimizing the use of health data, facilitating secure data sharing for secondary purposes such as research, innovation, policy-making, and regulatory activities. Building on the success of the MyHealth@EU infrastructure, which enables the exchange of e-prescriptions and patient summaries across 11 Member States, HealthData@EU represents a critical step forward in realizing the full potential of health data across the EU.

³³ <https://www.health-x.org/en/platform>

³⁴ <https://ehds2pilot.eu/>

The pilot project leverages the eDelivery Building Block services to facilitate the secure and interoperable cross-border use of health data. It designs, develops, deploys, and operates a network of nodes representing various data brokers, holders, and consumers supported by central services provided jointly by Member States and the European Commission. eDelivery AS4 Access Points ensures interoperability, while the eDelivery PKI service provides the necessary certificates to establish trust and security. The dynamic discovery model is used for managing participant information via the central eDelivery SML service, ensuring scalability and efficient data sharing.

3.3.2 Cognitive Ports data space

The future of the port industry lies in the designing and development of smart ports leveraging state-of-the-art technological innovations to create a framework that supports efficient and optimized operations. A smart port should be able to facilitate as many services and operations as possible, including traffic management, logistic operations, environmental conditions monitoring, inventory management, access and parking control, or operational safety and security. To this end, DataPorts ³⁵ aims to deliver an industrial data platform and intends to interconnect a cluster of heterogeneous digital infrastructures into a unified integrated ecosystem. The platform tries to establish the policies and rules for a reliable and trusted data-sharing framework, facilitating the establishment of a single data space for all European data ports and contributing to the EC global objective of creating a Common European Data Space. The proposed solution addresses the challenges of scalability, data source heterogeneity, and data governance mechanisms by introducing a three-layered architecture: (i) Data Access Component, which provides an interface to manage and integrate data into the platform, (ii) Data Processing Services, which provides semantic abstractions facilitating the functional connection between data sources and analytic services and (iii) Data Analytics and Cognitive Applications, which provide mechanisms to develop cognitive services supporting real-time insights from seaports' data.

3.3.3 Automotive data space

The automotive industry is a massive global sector that involves a wide range of operations related to manufacturing processes. Apart from the significant economic benefits to the world's economy, the automotive industry has raised a variety of international concerns, such as supply resiliency, sustainability, systematic coverage, and cost of innovation across the entire value chain. To address these concerns seamless collaboration among all automotive partners is imperative. In this direction, Catena-X ³⁶ is a collaborative network initiative aimed at establishing a secure and standardized data exchange infrastructure within the automotive industry. It seeks to enhance efficiency and transparency across the automotive value chain by facilitating seamless data exchange and collaboration among automotive manufacturers, suppliers, and other stakeholders. Catena-X has been implemented based on Eclipse Dataspace Connector (EDC), which acts as a central communication component within this ecosystem, providing a framework for sovereign, inter-organizational data exchange.

3.3.4 Mobility data space

Digitalization is a key enabler, making mobility smarter, sustainable, and more responsive to the needs of its users. Adapted with a sustainable and smart mobility strategy, it ensures the efficient movement of people, including tourists, commuters, business travelers, and other users, to, from, and within a destination. Therefore, the use of data has recently emerged as an essential approach for transport systems to establish an efficient, accessible, and sustainable ecosystem in which the data will be shared and exchanged in a reliable manner. A critical challenge for Europe is to enhance the potential of mobility data supporting the interoperability between existing and future transport and mobility data sources and ecosystems. Within this context, a common European mobility dataspace is the key technological point for the development of smart environments, in which the use of aggregated data will be offered in a single decisional platform. To this end, EONA-X ³⁷ aims to transform the Mobility, Transport, and Tourism (MMT) sector by establishing a secure and standardized data-sharing ecosystem. EONA-X is a dedicated European data space for Mobility, Transport and Tourism, developed on the path of the Gaia-X initiative.

³⁵ <https://dataports-project.eu/>

³⁶ <https://catena-x.net/en/>

³⁷ <https://eona-x.eu/>

3.3.5 Energy data space

The energy sector is crucial for the survival of all systems and infrastructures. Thus, Omega-X³⁸ is focusing on the development of a European Energy Data Space. This will offer a safe and reliable exchange of data that were previously unavailable, between regulators, energy companies, organizations, local community energy management systems, renewable sources grids operators, novel stakeholders such as aggregators, start-ups, and any other actor ready to provide and exploit datasets and information for the deployment of innovative services that will promote clean and sufficient energy distribution and smart energy grids management. The Omega-X is developed on the Gaia-X and IDSA principles of data exchange and its design is divided into the following main sectors: (i) the data and App Marketplace, which acts as the entry point for end-users in the data space ecosystem, through its graphical interface, (ii) the Federated infrastructure, which provides the mechanisms for secure and sovereign data exchanges, (iii) the connector, which enables data exchanges and service provision, and (iv) the Compliance Services, which fosters trust and interoperability.

3.3.6 Manufacture data space

The European manufacturing industry is a mature and influential sector that continues to play an important role in economic development. Consequently, the Smart Connected Supplier Network (SCSN) is an innovative initiative that aims to facilitate cross-factory communication, ensuring supply chain transparency and interoperability by reducing administrative burdens and fostering collaboration among supply chain partners. The SCSN prioritizes data sovereignty without the need for a centralized authority and operates based on the principles of interoperability and data sovereignty.

3.3.7 Earth observation data space

Copernicus is a European Union Earth Observation (EO) program that provides data and information on a wide range of environmental and security-related aspects. In this direction, the Copernicus Data Space Ecosystem³⁹ has been developed, involving various software components that manage, process, and disseminate the vast amount of data generated by the program. Some of the key components include:

- **Copernicus Services Information System (CAMS/C3S):** The Copernicus Atmosphere Monitoring Service (CAMS) and the Copernicus Climate Change Service (C3S) provide data, information, and tools related to atmospheric composition and climate change. These services have their own software components for data processing, modeling, and dissemination.
- **Copernicus Open Access Hub (SciHub):** SciHub is an online platform operated by the European Space Agency (ESA) that provides access to Sentinel satellite data, allowing users to search, browse, and download data from the Sentinel missions.
- **Copernicus Data and Information Access Services (DIAS):** These services provide access to a wide range of Copernicus data and information, including tools for data discovery, visualization, and download.
- **Copernicus Core Services Portals:** These portals provide access to specific thematic areas of Copernicus data and services, including (i) Copernicus Marine Environment Monitoring Service (CMEMS) Portal, (ii) Copernicus Land Monitoring Service (CLMS) Portal enhancing Forest Monitoring, and (iii) Copernicus Emergency Management Service (EMS) Portal.

In addition to the Sentinel satellites, the Copernicus program incorporates data from various contributing missions. These missions may have their own software components for data processing, storage, and dissemination. These software components work together to ensure that Copernicus data is efficiently managed, processed, and made available to users for various applications, including environmental monitoring, climate research, disaster management, and more.

³⁸ <https://omega-x.eu/>

³⁹ <https://dataspace.copernicus.eu/>

3.3.8 Agriculture data space

The Green Deal⁴⁰ and the Common Agricultural Policy ⁴¹ of the European Union impose immediate action on many different levels of political, social, financial, and technological integration across member states. On the technological level, the convergence of IoT networks and data spaces as data-generating and exchanging mechanisms for transforming agriculture, optimizing productivity, and ensuring sustainable practices. Accordingly, a common European agricultural data space has been created, supporting secure and trusted data exchange, enhancing economic and environmental performance. To this end, the AgriDataSpace ⁴² is a project that combines data space technologies (BDVA/IDSA/GAIA-X) with agricultural knowledge, new business models, and agri-environment policies. It relies on existing platforms and aims to introduce novel concepts and methods for creating sustainable innovation in the agriculture sector.

3.3.9 Finance data space

Data has always been at the core of financial services, encompassing individual savings, mortgages, consumer credit, investments, pensions and insurance, publicly disclosed company information, and business registry data. Therefore, to ensure interoperability, the EU Financial Data Space ⁴³ needs to be developed in close connection with data spaces in other related sectors. A common financial data framework aims to promote data-driven innovation in the finance sector, enabling (i) digital access to all publicly disclosed financial and sustainability-related information, (ii) easier reporting and sharing of supervisory data among EU and national supervisory authorities, and (iii) business-to-business and business-to-consumer data sharing and reuse in the EU financial sector.

3.4 Data space mechanism and connectors

3.4.1 Data space protocol

The Dataspace Protocol (DSP) is designed to facilitate secure, efficient, and standardized data exchange among various stakeholders within the European Data spaces.

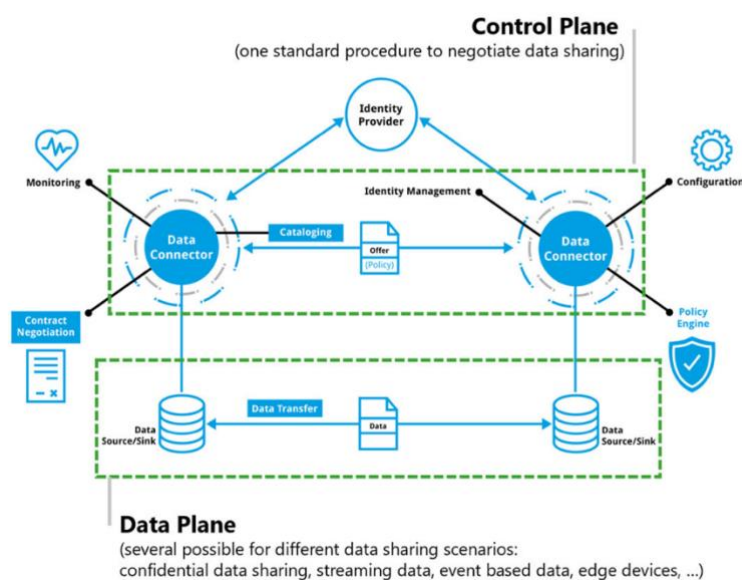


Figure 4: The Data space protocol as a technical innovation and strategic enabler ⁴⁴

⁴⁰ The European Green Deal - European Commission https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en

⁴¹ CAP at a glance - European Commission https://agriculture.ec.europa.eu/common-agricultural-policy/cap-overview/cap-glance_en

⁴² <https://agridataspace-csa.eu/>

⁴³ <https://www.european-financial-data-space.com/>

⁴⁴ <https://internationaldataspaces.org/offers/dataspace-protocol-overview/>

It addresses the challenges of data interoperability, data sovereignty, and secure data sharing, providing a common framework that ensures minimum viable interoperability for systems and organizations that wish to exchange data seamlessly. The DSP is envisioned as a strategic enabler for data-driven economies and an impactful specifications protocol similar to protocols that shaped the internet and the communication industries (Figure 4). The DSP orchestrates all necessary steps for parties to share their data, including requesting a catalog, negotiating a contract, and managing the transfer process.

Specifically, to facilitate interoperable data sharing, the provision of metadata is required. The Data space protocol defines how these metadata are provisioned by identifying:

1. The structure of the datasets and the associated usage control policies.
2. How contract agreements that govern data usage are expressed and negotiated.
3. How the data transfer process is enforced (via transfer process protocols).

The latest version of the Data space Protocol is the 2024-1 version (released in March 2024), and the upcoming Data space Protocol Specification is scheduled for release in late summer this year.

3.4.2 Data space components

3.4.2.1 Identity management

On a conceptual basis, identity management pertains to the identification, authentication, and authorization of stakeholders operating in a data space. It ensures that organizations, individuals, machines, and other actors are provided with acknowledged identities and that those identities can be authenticated and verified, including additional information provisioning, by authorization mechanisms to enable access and usage control.

Throughout the development of data spaces, identity management as a set of functionalities has been realized by a number of components, ranging from the Fraunhofer's Identity Provider⁴⁵ component to the Identity Hub⁴⁶ and Registration Service⁴⁷ components, implemented by the Eclipse Foundation⁴⁸, while maintaining its core conceptual rigor as a trusted technology system that creates, maintains, and manages identity information for a data space participant.

3.4.2.2 Broker

Within a data space, the Broker acts as a central registry, storing and managing information about data sources and the available assets. It provides an interface for data providers to submit metadata about their offerings, while ensuring search capabilities for data consumers to locate relevant recourses. In EDC implementations, the Broker's functionality is realized by a Federated Catalog Service, which constitutes an indexed repository of Self-Descriptions to enable the discovery and selection of data providers and their service offerings. The Federated Catalog can retrieve all contract offers by every provider connector participating in the data space, namely all assets bound with a usage policy within a contract definition.

3.4.3 Data space connectors

This section brings clarity to the implementations of data connectors. Firstly, in Figure 5 we provide an overview of available data space connectors in the community based on their characteristic. Each connector is described according to a specific structure: (i) connector overview (e.g., connector name, maintainer), (ii) license type (i.e., open source, open source – copyleft, closed source – extendable, closed source), (iii) IDS certification, (iv) identity management (e.g., centralized/X.509, decentralized/did:web, decentralized/SSI, None), (v) deployment options (e.g., edge, on-premises, cloud, IoT/CPS/OT devices), and (vi) service level (e.g., connector-as-a-service, platform-as-a-service, self-service). Secondly, we provide a technical overview of the most popular and relevant for the project needs elected, which have been thoroughly investigated in the context of the NextGEM project.

⁴⁵ <https://github.com/International-Data-Spaces-Association/omejdn-daps>

⁴⁶ <https://github.com/eclipse-edc/IdentityHub>

⁴⁷ <https://github.com/eclipse-edc/RegistrationService>

⁴⁸ <https://www.eclipse.org/org/foundation/>



















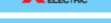















Connectors with up-to-date information							
Section	Name of connector	Maintainer	Open source	IDS certified	Identity management	Deployment options	Service level
2.1.1	Advaneo Open-Source EDC Connector		✓		X.509	• On-premises • Cloud	Self-service
2.1.2	Advaneo-X Connector				X.509	• On-premises • Cloud	• Connector-as-a-Service • Platform-as-a-Service • Self-service
2.1.3	Advaneo Trusted Connector				X.509	Cloud	• Connector-as-a-Service • Platform-as-a-Service
2.1.4	ALSOV Connector					• Edge • On-premises • Cloud	Platform-as-a-Service
2.1.5	Boot-X Connector						Connector-as-a-Service
2.1.6	Data Space Integration				• X.509 • didweb • SSI	Cloud	Platform-as-a-Service
2.1.7	ECI IDS Connector powered by TNO				X.509	Cloud	• Connector-as-a-Service • Self-service
2.1.8	Eclipse Dataspace Components (EDC)		✓			Not specified	Service level is best effort of the open-source community
2.1.9	EdgeDS Connector		✓			• Edge • On-premises • Cloud • IoT/CPaaS/OT	Connector-as-a-Service
2.1.10	EGI Datahub Connector		✓		X.509	• On-premises • Cloud	Platform-as-a-Service
2.1.11	EOANA-X EDC Connector		✓		didweb	• On-premises • Cloud	• Connector-as-a-Service • Self-service
2.1.12	FIWARE Data Space Connector		✓		X.509	• On-premises • Cloud	• Connector-as-a-Service • Self-service
2.1.13	GATE Dataspace Connector		✓	✓	X.509	On-premises	Platform-as-a-Service
2.1.14	GDSO Connector – Tyre Information Service		Partially		AIRIS Cogito	• Edge • On-premises • Cloud • IoT/CPaaS/OT	Platform-as-a-Service
2.1.15	HEALTH-X dataLOFT EDC				SSI	• On-premises • Cloud	• Connector-as-a-Service • Self-service
2.1.16	IIOC (Intel IONOS Orbiter Connector)		✓		SSI	• Cloud • IoT/CPaaS/OT	
2.1.17	Kharon IDS Connector				Kharon	• Edge • Cloud	• Connector-as-a-Service • Platform-as-a-Service
2.1.18	Mitsubishi Dataspace Connector		Partially			IoT/CPaaS/OT	Self-service
2.1.19	MPAD-C by Mondragon				X.509	On-premises	Self-service
2.1.20	OneNet Connector		✓		X.509	On-premises	Connector-as-a-Service
2.1.21	Prometheus-X Dataspace Connector		✓		didweb	• On-premises • Cloud	• Connector-as-a-Service • Platform-as-a-Service
2.1.22	Silicon Economy EDC		✓			• Edge • On-premises • Cloud • IoT/CPaaS/OT	Self-service
2.1.23	sovit CaaS (Connector-as-a-Service)				• X.509 • didweb • SSI	• On-premises • Cloud • Others	Connector-as-a-Service
2.1.24	sovit Open-Source EDC Connector		✓		• X.509 • Mock IAM	• On-premises • Cloud • Others	Self-hosted
2.1.25	TANGO Connector		Partially		• didweb • SSI	• On-premises • Cloud	Platform-as-a-Service
2.1.26	Tech2B SCSN Connector				didweb	Cloud	
2.1.27	Tekniker Dataspace Connector		✓		SSI	• On-premises • Cloud • Edge	Connector-as-a-Service
2.1.28	Telekom DIH Connector			✓	• X.509 • didweb	• On-premises • Cloud	Connector-as-a-Service
2.1.29	TNO Security Gateway (TSG)		✓	✓	X.509	Cloud	Self-service
2.1.30	Triton Enterprise Connector					• On-premises • Cloud	Platform-as-a-Service
2.1.31	TRUE Connector		✓	✓	X.509	• Edge • On-premises • Cloud • IoT/CPaaS/OT	• Connector-as-a-Service • Platform-as-a-Service • Self-service
2.1.32	Trusted Connector		✓		X.509	• Edge • On-premises • Cloud	Platform-as-a-Service
2.1.33	Trusted Supplier Connector (TSC)					• Edge • On-premises • Cloud	• Connector-as-a-Service • Platform-as-a-Service • Self-service
2.1.34	VIT DSIL Connector			✓	X.509	• On-premises • Cloud	• Connector-as-a-Service • Platform-as-a-Service

Figure 5: An overview of data space connectors (IDSA) ⁴⁹

⁴⁹ https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-Data-Connector-Report-84-No-16-September-2024-1.pdf

3.4.3.1 IDS Connector

The IDS Connector is an implementation of an IDS connector component, which follows the IDS Reference Architecture Model. It integrates the IDS Information Model⁵⁰ and uses the IDS Messaging Services⁵¹ for IDS functionalities and message handling. The project was supported and promoted by the International Data Spaces Association (IDSA), but it is not actively maintained. The core component of IDS Connector provides REST APIs for loading, updating, and deleting resources with local or remote data enriched by their metadata. Figure 6 illustrates the data model, consisting of the following entities. The Resources (i.e., namely metadata records) are organized in Catalogs and are associated with Representations, where the offered data are described in more detail. Each Representation is connected to the Artifact entity, which has a 1:1 relationship with the raw data and where the URL indicating the location of the data is provided. Resources are also associated with Contracts, and each Contract can contain multiple Rules that describe the Usage Control patterns pertaining to the specific Resource.

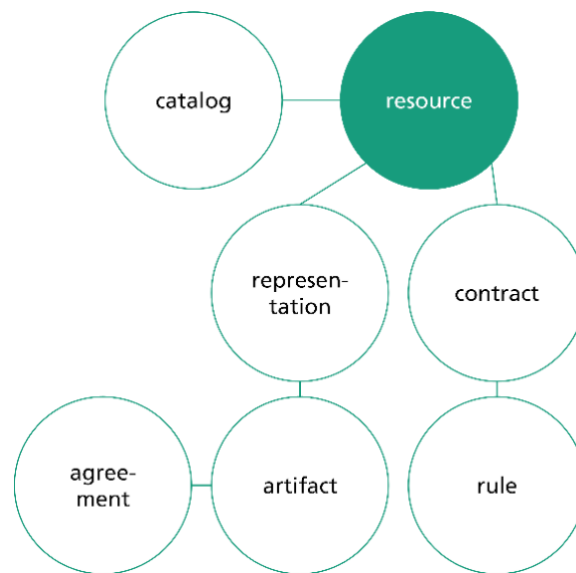


Figure 6: IDS Connector data model ⁵²

With regards to the Usage Control policies defined, there are 21 rules described, 8 of which are supported by the Data space Connector implementation, including the access without restrictions (public), the connector restricted access, where only a specific connector can obtain the data and the time restricted access that allows for the data to be exchanged within a specific time interval.

3.4.3.2 Eclipse data space connector

The Eclipse Dataspace Connector (EDC) is an open-source project hosted by the Eclipse Foundation. It comprises an IDS-compliant, modular and highly extensible framework that can support alternative protocols, where contrary to previous implementations such as the IDS Connector, it provides a separation of the control and data plane, allowing both planes to operate independently without depending on one other.

The core modules contain all the methods, model classes and interfaces, which are necessary for building and running a connector with the minimum capabilities, whereas the extensions are implemented as Java plugins that extend the connector's core functionalities. This extensibility adds new features and/or extends the core functionalities in terms of transfer processes, metadata storage solutions, identity management components and supported data exchange policies. The extensions are implemented using EDC Service Provider Interface (SPI), which provides a standardized way for developers to create custom extensions. Figure 7 illustrates conceptually the EDC architecture and its extensions' mechanism. The main extensions are summarized as follows:

- **Data Plane HTTP Extension.** This extension supports the integration of the data plane with HTTP object storage, enabling seamless data transfer between EDC data plane and external HTTP-based storage services, providing data storage with a robust and scalable manner.

⁵⁰ <https://github.com/International-Data-Spaces-Association/InformationModel>

⁵¹ <https://github.com/International-Data-Spaces-Association/IDS-Messaging-Services>

⁵² <https://international-data-spaces-association.github.io/DataspaceConnector/Documentation/v6/DataModel>

- **Federated Catalog Extension.** This extension enhances the EDC with the ability to support federated catalog services, enabling each connector to issue requests to other catalogue services and gather the results into a single unified view. This unified view facilitates metadata access across different providers, improving data discovery processes.
- **Basic Policy Functions for Data Consumption.** This extension enhances the EDC by introducing policy functions that enable the validation of data consumption, enabling data owners to establish temporal constraints on data access, and ensuring that data is consumed based on specific predefined policies.
- **Decentralized Identifier Extension.** This extension provides the mechanisms required to enable the EDC to resolve DIDs, and verify identities using DIDs and Verifiable Credentials.

The data model of the EDC includes:

- **Asset.** It represents the data or service that is being shared within the data space through EDC, defining the unit of sharing.
- **Policy.** It represents a collection of rules that govern usage permissions or restrictions in data sharing.
- **Contract Offer.** It represents specific obligations and permissions associated with an Asset and acts as a starting point for the negotiation process.
- **Contract Agreement.** It represents a formal agreement between parties and results from a contract negotiation process.

The Contract Offers are visible in the Catalog that connectors can access in the form of a provider's Contract Offers. Upon successful Contract Negotiation and evaluation of the respective Policy, a Contract Agreement is created between the provider and consumer parties, and the transfer can then take place.

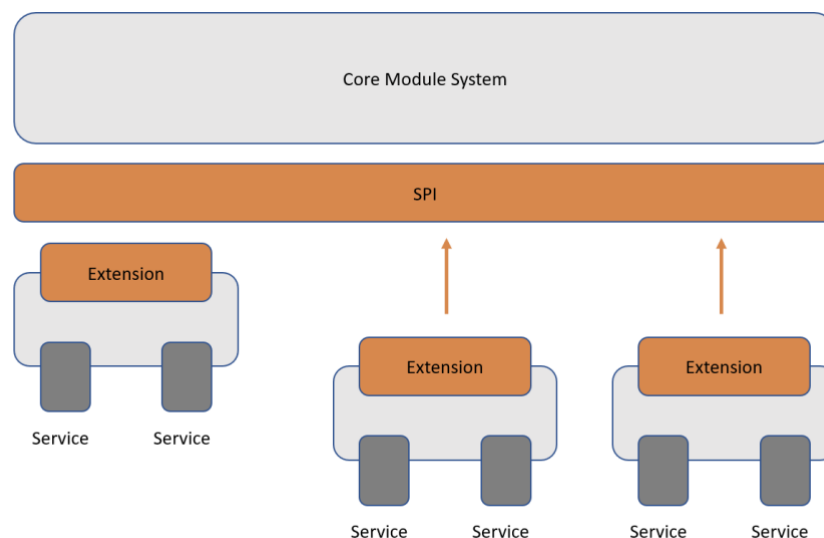


Figure 7: EDC Core Module System and its architecture

3.5 Initial deployment of data space solutions

One of the most important steps in developing software products is the definition of the Minimum Viable Product (MVP), which is needed to start the development process by integrating the essential features that meet early adopters' needs. The main objective of an MVP is to validate the core concept of the product while minimizing the time, resources, and costs involved in its development. An MVP focuses on the most critical features and functionalities that address the primary problem that the product aims to solve. Building on the concept of an MVP, various organizations involved in the design and implementation of data space solutions have introduced the Minimum Viable Data space (MVD), a minimal deployment that focuses on the core functionalities required to establish a functional data exchange environment. This section presents each of these minimal configurations as the primary deployments for designing and implementing the NextGEM platform. The various solutions were tested in terms of data types supported for exchange, how the file exchange is implemented, if an identity management mechanism is supported, as well as which policies that are relevant to the NIKH platform are available. In the following subsections, the validation of each of the implementations is described and a comparative table focusing on the aforementioned features is provided.

3.5.1 Single-connectors MVD

The Samples EDC repository ⁵³ consists of a collection of scenarios that support the basic functionalities of configuring and running an EDC connector, as well as the steps for completing various data exchange examples. The Samples are organized into Basic and Transfer scopes, and recently, the Advanced and Policy scopes have been added. In all of the scenarios in the Samples repository, the connectors are built with Gradle Wrapper, using Gradle build files that contain the necessary dependencies.

The basic scenarios focus on explaining the procedures of running and configuring a simple connector and writing extensions to add features and functionalities to the Connector. The examples under the Transfer scope are related to the steps required for building and running a provider and a consumer Connector, creating Assets, Policies and Contract Definitions on the provider's side, as well as accessing the catalog, initiating contract negotiation and requesting data transfer on the consumer's side.

The Transfer scenarios are separated into local data transfer and cloud data transfer cases and are further divided into the consumer-pull and provider-push scenarios. The consumer-pull scenario corresponds to the case, in which the data provider exposes a REST API serving data and the consumer is interested in continuously querying this API with different parameters, to obtain the data they need, without the need to establish a new contract agreement every time the consumer makes a new query, whereas in the provider-push case, the consumer wants to perform a one-time transfer of a data object. The latter case is relevant to the NIKH data space services, where data are exchanged between parties in an ad-hoc manner, and contracts are necessary to ensure sovereignty over every exchanged asset.

Figure 8 illustrates a provider-push data transfer scenario, in which the data is exchanged between a provider and a consumer connector in the form of an HTTP URL, containing a JSON object (i.e., the data defined in the Asset creation endpoint are of the type "HttpData"). The consumer must run an HTTP logging server that logs all incoming requests, enabling access to the transferred data. The address of the logging server is specified in the transfer process endpoint, and once the transfer process is completed, the JSON object of the exchanged URL can be seen in the logs of the backend service.

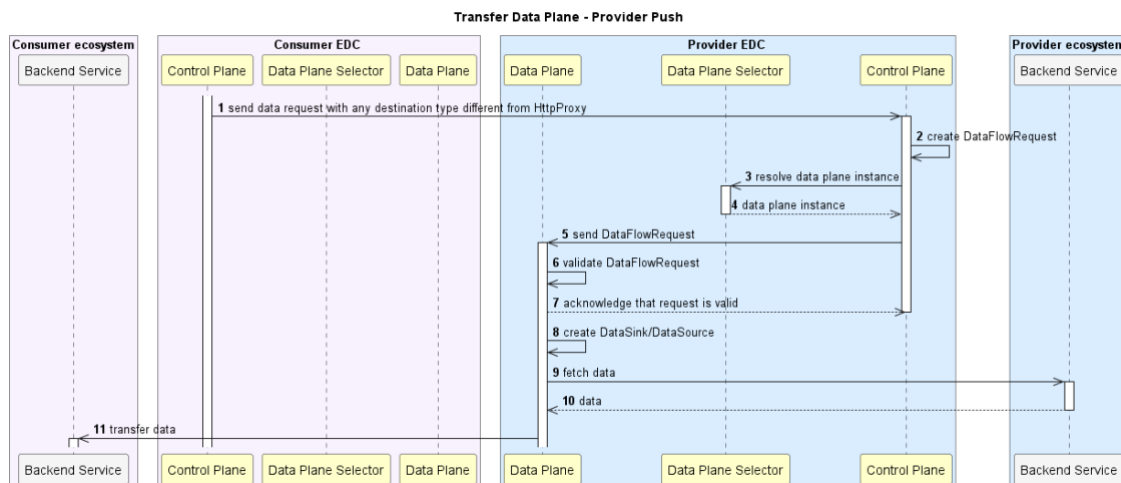


Figure 8: Sequence diagram of the Provider Push paradigm

3.5.1.1 Single connectors MVD validation.

The Samples EDC repository has been a helpful implementation for originally exploring and understanding the EDC framework. Apart from the Basic scenarios, the samples that were explored in depth were the Transfer samples related to building and running a provider and a consumer connector (Transfer Sample 00), creating Assets, Policies and Contract Definitions on the provider's side, and initiating the negotiation on the consumer's side (Transfer Sample 01), as well as performing a provider-push data exchange scenario by completing a file transfer (Transfer Sample 03).

On the connectors' side, the states of the Transfer Process are logged both for the provider and the consumer. On the provider's side, a successful Transfer Process begins in state INITIAL, moves on to state PROVISIONING and then PROVISIONED, and afterwards from state STARTING to STARTED, until it

⁵³ <https://github.com/eclipse-edc/Samples>

reaches the COMPLETED, DEPROVISIONING and DEPROVISIONED states. On the consumer's side, the respective Transfer Process begins from state INITIAL, changes into states PROVISIONING and PROVISIONED, and moves on to REQUESTING and REQUESTED states, until it reaches the COMPLETED state.

The initial testing was performed via a centralized local deployment, with the goal of familiarization with the EDC APIs. Additionally, in the local deployment, efforts were made to extend the Samples' functionalities to support file exchange in the local filesystem. In the meantime, the two connectors of the examined Samples were also deployed into separate host machines in order to support the distributed paradigm, as envisioned in the NIKH platform. The connectors were running as containers, deployed in Kubernetes, and their addresses were exposed to establish the communication between them. The provider and consumer could successfully exchange JSON data, which were logged in the HTTP logging server; however, the Samples' unsupported functionalities in terms of actual file exchange (not "HttpData") and in terms of identity management highlighted the need for a more complete connectors' implementation.

3.5.2 IDSA MVD

The IDSA MVD is a combination of components that initiate a data space with just enough features to be usable for secure and sovereign data exchange, as specified by IDS Reference Architecture. It is based on IDS Connector and provides the IDS Deployment Examples that were utilized for understating the appropriate APIs for uploading resources (metadata records) on the provider's side, defining rules for the usage and access control of resources, negotiating contracts, and initiating the file exchange process.

The IDS Connector was locally deployed, and the provided Swagger-UI was used for exploring and testing various scenarios of resource offering and usage control rules imposition, whereas in the IDS Deployment Examples, apart from the Swagger-UI, predefined scenarios were provided in the form of scripts. Apart from the provider-consumer IDS Deployment Example that was locally deployed, the Kubernetes deployment of the slim IDS Deployment Example via Helm was also tested.

In the IDS Connector, the supported files are given in the form of text or JSON files via a URL and after a successful file transfer, the consumer has the option to download the content of this URL in the form of a file. As far as the tested rules are concerned, the "Provide Access", "Prohibit Access", and "N Times Usage" policies were explored, where in the latter, it was validated that the consumer could download the requested file only as many times as defined in the respective rule accompanying the offered resource. Efforts were also made to incorporate and test additional components for identity management, such as the DAPS (Dynamic Access Provisioning Service) and the Certificate Authority components. However, the efforts were not successful.

The Data space Connector stopped being actively maintained, and subsequently, the connector implementations exploration shifted towards the EDC connector in order to discover the options with the most suitable functionalities to address the needs of the NIKH platform Data space Services.

3.5.2.1 IDSA MVD validation

The main project repository can be used to locally run and test a minimum deployment by creating two connectors representing the data provider and consumer, respectively. Through the provided interactive Swagger-UI, Resources and their respective Catalogs, Artifacts, Contracts, Representations, and Rules can be created on the provider's side to fully describe a registered Resource. Additionally, data exchange processes can be initiated on the consumer's side, by first requesting the provider Connector's self-description to view the offered resources, requesting the metadata of an offered resource, including the contract offer, and then negotiating the Contract, to be in accordance with the rules the provider has defined, and reach an Agreement. Then, on the consumer's side, the data can be downloaded in the form of a file containing the data that was initially given through the URL in the provider's Artifact.

Apart from the main project repository, the IDS Deployment Examples repository is available and contains three different setups, namely a full setup including the Data space Connector UI and a Postgres database to store the metadata information, a simpler provider-consumer setup with a predefined data exchange example between a provider and a consumer, without a UI and a slim setup with minimum functionalities.

The main project repository enables to build and run the connectors locally either through a step-by-step guide using Maven, or via a Docker image, while the IDS Deployment Examples repository offers the possibility of running all the setups by deploying the connectors via Docker, but especially for the slim setup the option of a Kubernetes deployment is also provided.

3.5.3 Sovity MVD

The Sovity Minimum Viable Data space (SMVD) is a demonstration deployment that showcases the capabilities of the Eclipse Data space Components (EDC)⁵⁴. Based on the principle that the EDC is a highly modular and extensible framework, an open-source deployment by Sovity has been created by expanding its functionalities through the concept of Extensions. The project by Sovity provides a list of Docker images with varying features and scopes, ranging from an open-source development edition with a minimum set of features and the purpose of manual testing to a commercial edition for use in production being offered in a Connector-as-a-Service paradigm.

The setup of the development edition includes two connectors, their respective UI services, and two Postgres services for storing the metadata of created Assets, Policies, Contracts, etc. The transfer scenario that is supported by this implementation is similar to the one in the Samples project, using a URL containing a JSON object. Once the negotiation is completed successfully and the transfer is initiated by the consumer connector, the transfer is marked as completed, but no actual exchange is performed, since this functionality is not supported by this implementation. Additionally, the identity management feature is also omitted in the development edition of this connector, and therefore no authentication mechanism is utilized in this case.

The additional extensions supported by the Sovity connector in the form of extensions, mostly relate to a modified UI, based on the original Data Dashboard project by the Eclipse Connector, as well as policies that are related to giving restricted access to a specific connector and defining a time interval, during which the access to an Asset is allowed.

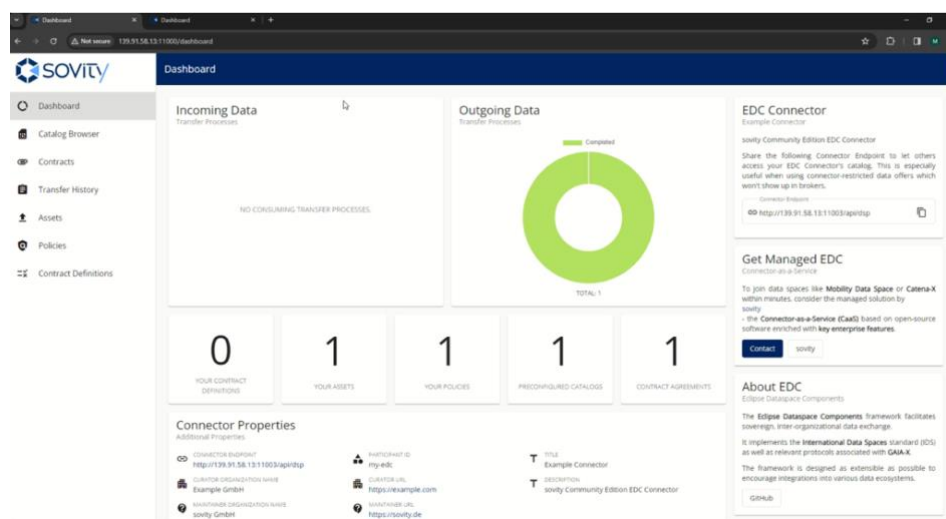


Figure 9: Sovity MVD UI Dashboard

3.5.3.1 Sovity MVD validation

The Sovity MVD offers Docker images for the EDC with some extensions, as well as a user-friendly interface that facilitates the interactions with the data connectors. Through the UI several scenarios of Asset, Policy and Contract Definition creations, as well as examples of accessing the Catalog, negotiating Contracts and viewing the Transfer History were tested. The Dashboard main page can be seen in Figure 9.

For the Asset creation, the supported fields in the UI only allow the user to input a URL as the actual location (baseUrl) of the data, as in the EDC Samples. In terms of Policies, additional ones are supported, namely the Time-Period-Restricted and the Connector-Restricted policies. The first one enables the provider to define a period during which access to their data is allowed, which matches the Public After Embargo privacy level of the NIKH metadata records, while the second one corresponds to the Restricted privacy level since it is meant to grant access to the provider's data for a specific data space participant. The creation of a Time-Period-Restricted policy through the Sovity UI can be seen in Figure 10. In addition, Figure 11 depicts the contract negotiation process, while Figure 12 shows the transfer history for a provider.

The Sovity deployment was first tested in a local deployment, where the functionalities of the additional Policies and the UI were examined. By running an additional connector to the initial provider-consumer setup, it was validated that when two of the connectors offer data, the Catalog of Contract Offers is updated with all of the

⁵⁴ <https://github.com/eclipse-edc>

providers' offers, as in the MVD. Furthermore, efforts were made to support the file exchange functionality, which is absent in this implementation. Meanwhile, the Sovity connectors were also deployed in a distributed manner in two separate host machines, and their addresses were exposed to enable the communication between them in order to facilitate the NIKH platform distributed data space setup.

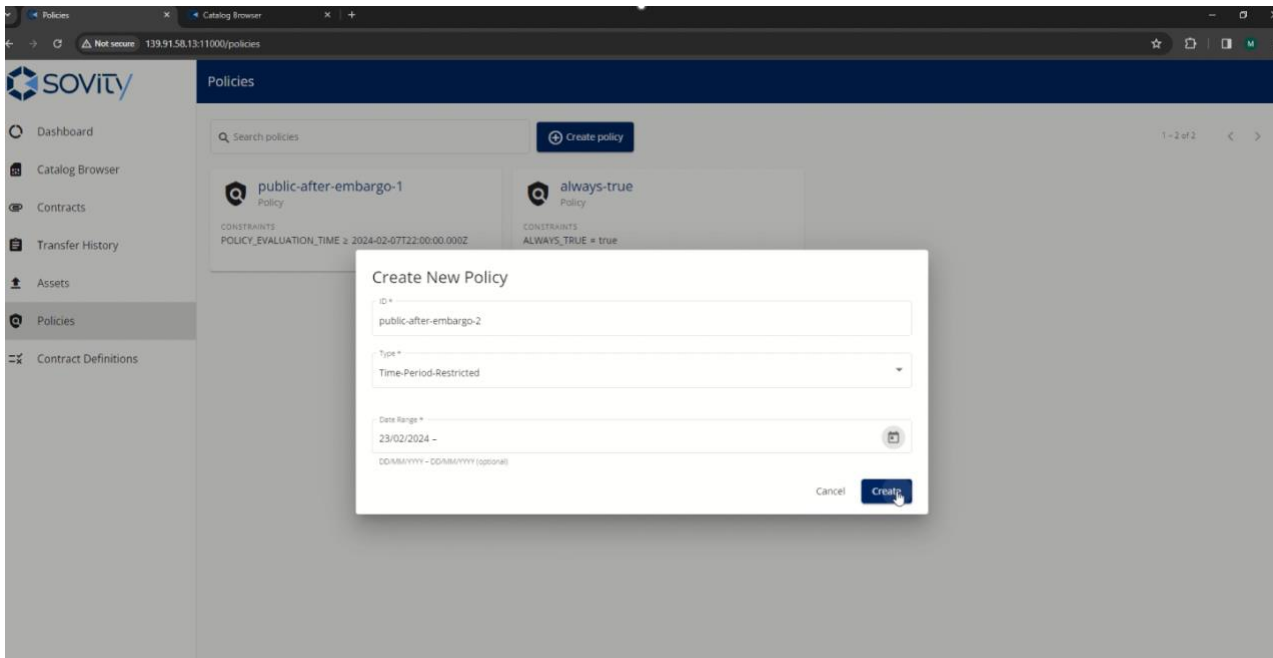


Figure 10: Time-Period-Restricted Policy Creation in the Sovity Connector UI

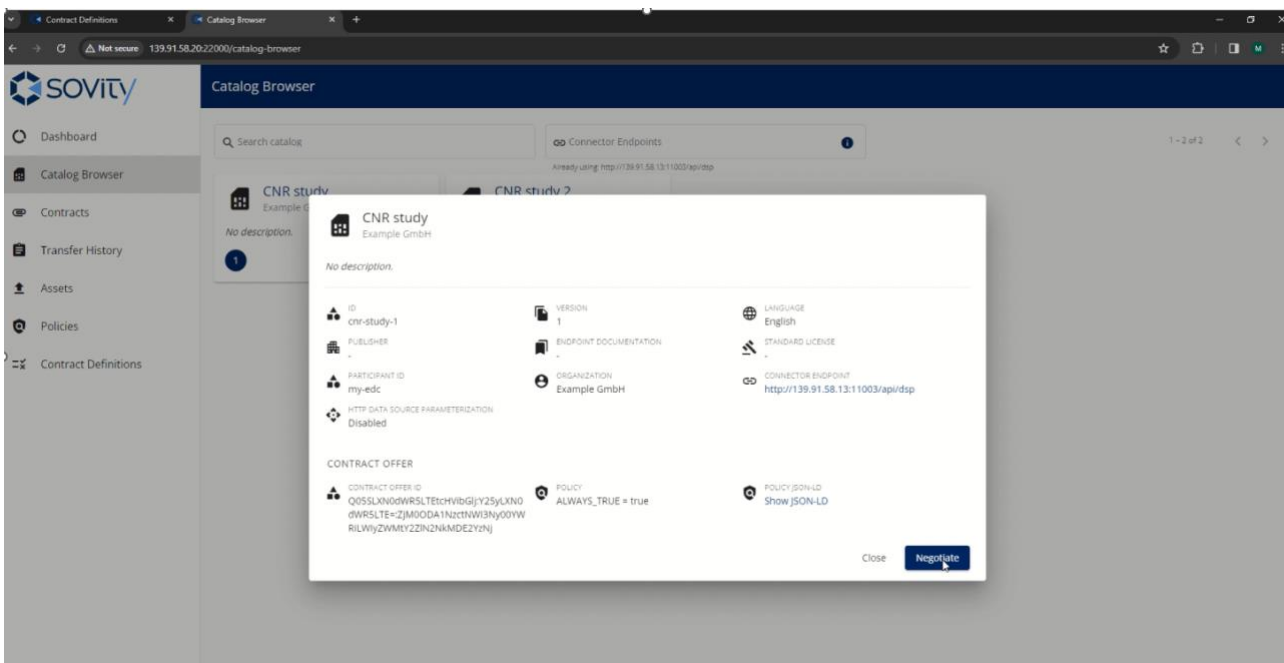
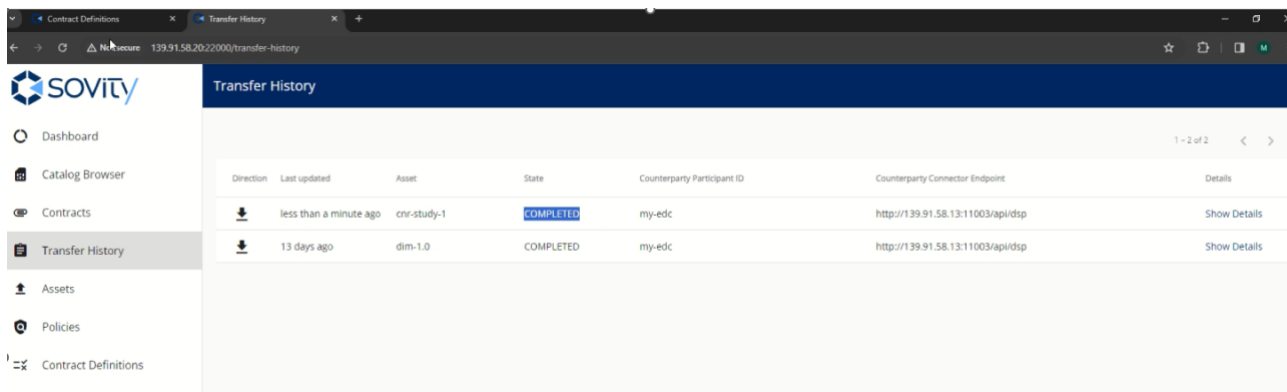


Figure 11: Contract Negotiation in the Sovity Connector UI

However, the lack of the actual data exchange functionality and the absence of any Identity Management mechanism indicated the need to use another more complete implementation. Meanwhile, the additional policies that exist as extensions in the Sovity connector and perfectly match the NIKH data privacy levels will be leveraged in the form of extensions in the MVD implementation to cover the needs of the NIKH Data space Services.



Direction	Last updated	Asset	State	Counterparty Participant ID	Counterparty Connector Endpoint	Details
↓	less than a minute ago	cnr-study-1	COMPLETED	my-edc	http://139.91.58.13:11003/api/dsp	Show Details
↓	13 days ago	dim-1.0	COMPLETED	my-edc	http://139.91.58.13:11003/api/dsp	Show Details

Figure 12: Transfer History list in the Sovity Connector UI

3.5.4 Eclipse MVD

The Eclipse Minimum Viable Data space (EMVD) is a demonstration deployment that showcases the capabilities of the Eclipse Data space Components (EDC) ⁵⁵. The EDC project is an open-source project under the governance of the Eclipse Foundation ⁵⁶. It provides a concrete example of how abstract the data space concepts can be implemented in real-world scenarios and demonstrates a comprehensive overview of the features designed and developed to support the Eclipse Data space Connector ⁵⁷ in full compliance with IDSA requirements on IDS protocol, rules and agreements of the RAM, and compliance with the IDSA certification scheme. Figure 13 presents the components that compose this initial deployment.

- **Connector.** The connector is based on Eclipse Data space Connector and represents the organizations that simulate data providers/consumers in different locations (e.g., EU, US). The Connector allows data owners and data providers to exchange and share their data with other participants and contains modules for performing data queries, data exchange, policy enforcement, monitoring, and auditing.
- **Azurite.** This service is based on an Amazon storage emulator ⁵⁸ and provides object storage capabilities to connectors. It is essentially a local emulator that simulates the functionalities of Azure Blob Storage on MVD's deployment.
- **Registration Service.** The Registration Service in the EDC architecture plays a key role in managing participants within a data space, facilitating the on-boarding process of new participants, and providing a catalog for all participants in the data space.
- **Data Dashboard.** This service is the input data component to the connectors, as well as the retrieval component responsible for acquiring and presenting asset-related data to the users. The main functionalities include: (i) asset creation, (ii) asset discovery, and (iii) policy and contract definition.
- **DID Server.** This service serves participants' decentralized identifiers.
- **CLI Service.** This service provides a CLI interface facilitating the interactions with the data space services.
- **Initialization Service.** This Service uses a postman image to run data seeding scripts providing automation in connectors' deployment. These scripts handle (i) asset creation, (ii) policy creation, and (iii) contract definition by associating assets and policies.

⁵⁵ <https://github.com/eclipse-edc>

⁵⁶ <https://www.eclipse.org/org/foundation/>

⁵⁷ <https://github.com/eclipse-edc/Connector>

⁵⁸ <https://github.com/Azure/Azurite>

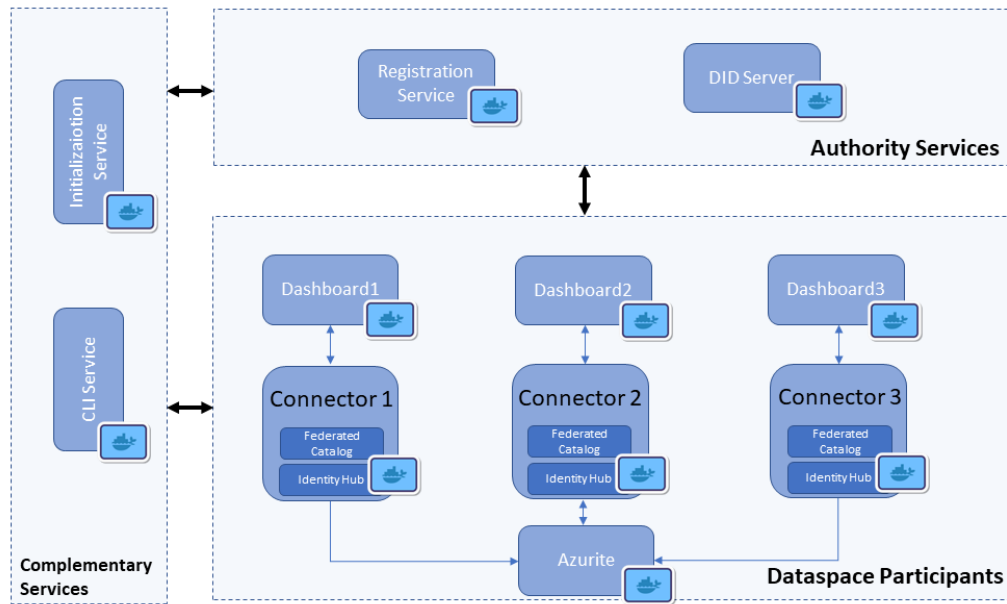


Figure 13: Eclipse Minimum Viable Data space deployment

The scenarios covered by this deployment are the creation of Assets, Policies, the viewing of the Catalog, the Contract negotiation, the data transfer initialization, and the browsing of Contract Agreements and Transfer history. An important part of the Eclipse MVD is the aforementioned Data space Authority, which in this setup is centralized, but the aim is to be decentralized in the future. Its functionalities include the issue of Verifiable Credentials, the establishment of DIDs (Decentralized Identifiers) for the participants, the enrolment process of the data space participants that are allowed to be part of the data space, and the listing of their DIDs. The last two functionalities are supported by the Registration Service, which is part of the Data space Authority. The Registration Service is also responsible for seeding the Catalog with all of the participants' contract offers, therefore making it a federated catalog. With regard to the DIDs, they represent a decentralized digital identity and are used to authenticate IDS calls between participants of the MVD.

The flow sequence for the distributed authorization among the data space participants is described in Figure 14. In the case described in the diagram, in order for Participant A to access a service provided by Participant B, their identity has to be established. The DID of Participant A is sent along with a bearer token to Participant B, allowing Participant B to authenticate the request by validating the JWS signature against the public key in the DID document. Then, Participant B obtains the Verifiable Credentials of Participant A from their Identity Hub and validates it against the public key in the DID document. Lastly, the access policy for the requested service is applied and a response is sent to Participant A authorizing or rejecting the request.

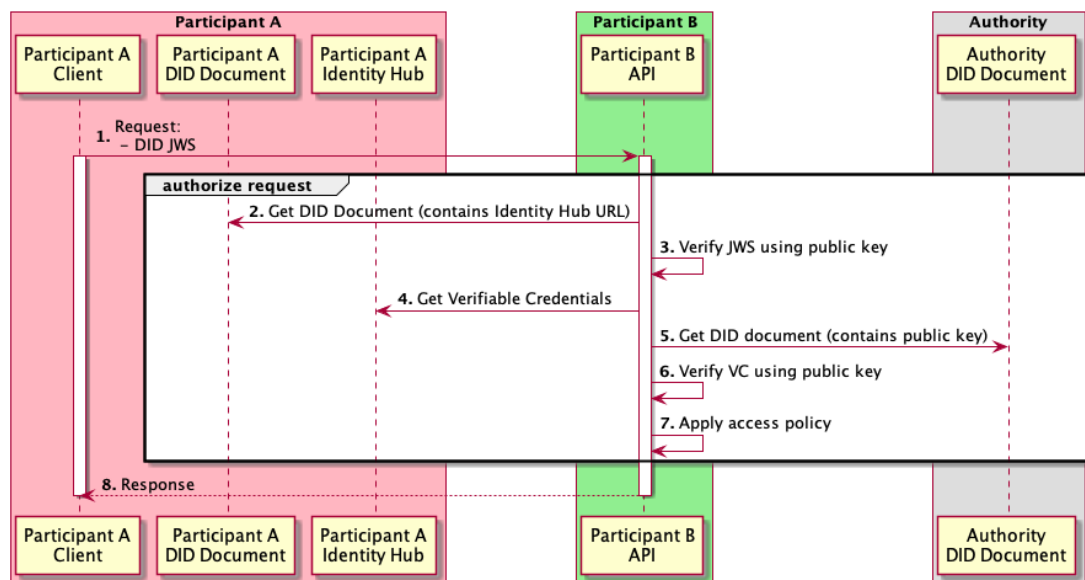


Figure 14: Flow sequence diagram of distributed authorization between data space participants in MVD

There are two options supported for the deployment of the Eclipse MVD, the local deployment setup and the cloud environment setup. Since the cloud setup requires a subscription to Azure, only the local deployment has been tested in the NextGEM project. With regard to the transfer process supported by the local MVD deployment, the participants (connectors) can exchange data by creating an Azure storage account and connecting to the local blob storage account through Azurite. The connection can either be established with CLI commands or by using the Microsoft Azure Storage Explorer desktop application. To exchange data, participants have to create a container in their storage account, in which the data file to be exchanged via the connectors has to be uploaded.

3.5.4.1 Eclipse MVD validation

In order to validate the Eclipse MVD, we tested an example that demonstrates the end-to-end capabilities of this deployment, investigating the file transfer between two participants in a decentralized data space. This minimum data space offers a basic deployment, enabling specific scenarios, as well as a user-friendly interface that facilitates the interactions with the data connectors. We set a local development environment that involves running Docker containers to simulate the three different participants in the data space. The user interface provides a comprehensive suite for (i) creating and managing assets, contract definitions and policies, (b) viewing the catalog browser, which lists the contract offers provided by the other participants, and (iii) viewing the transfer history. Figure 15 shows the main page of the dashboard, highlighting the catalog that lists all available assets offered by other participants within the data space.

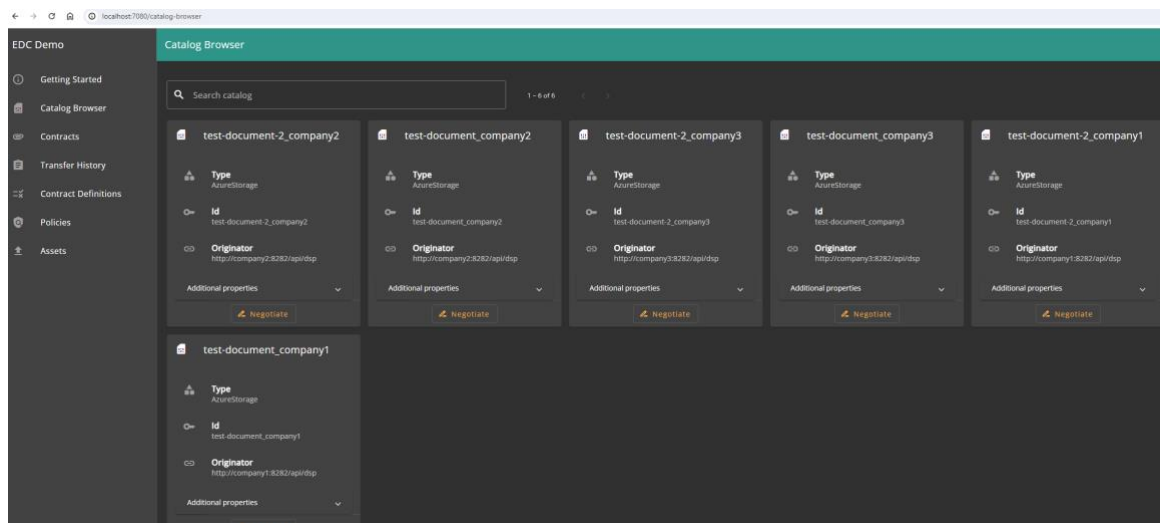


Figure 15: Catalog browsing lists all the available assets within the data space

In order to associate an actual dataset with an asset, we connected to a local storage account of the first participant using the Microsoft Azure Storage Explorer for (i) creating a container named “src-container” and (ii) uploading a file named “test-document.txt” to this container as illustrated in Figure 16. The uploaded actual dataset has been associated with an asset named “test-document_company1” and the appropriate policy as depicted in Figure 17, through Contract Definitions section.

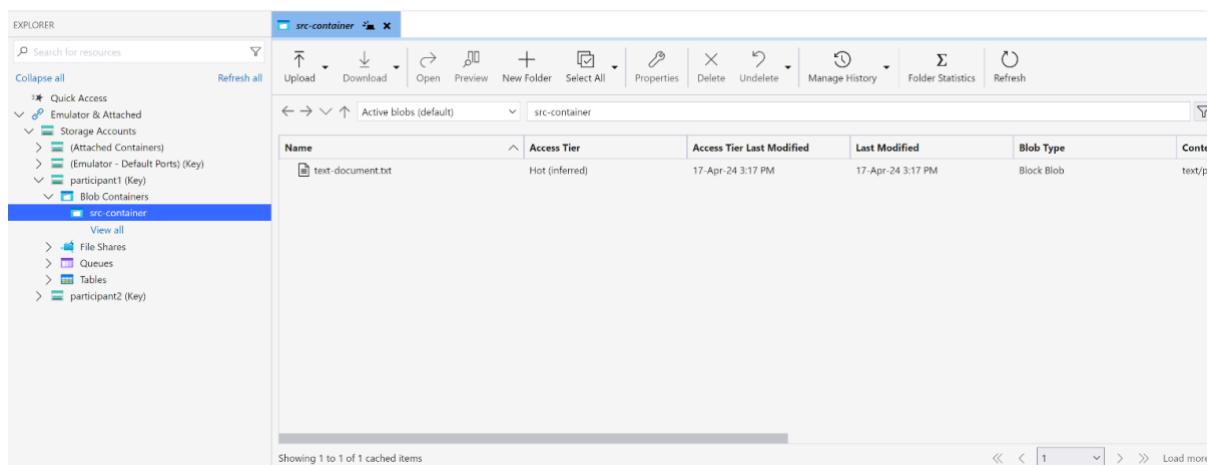


Figure 16: Blob storage and its contents

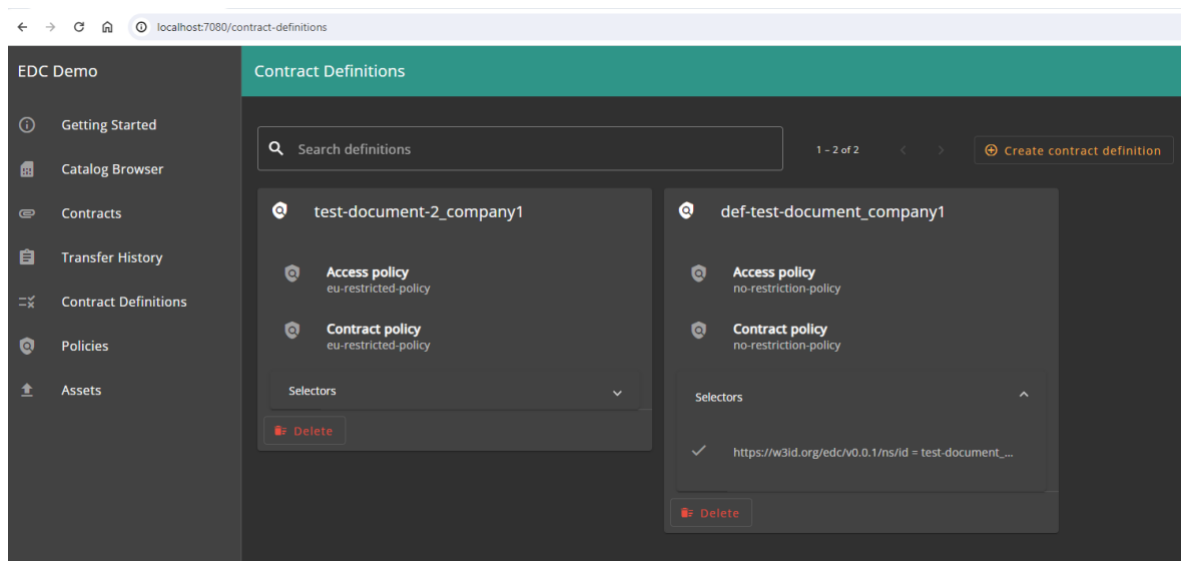


Figure 17: A contract definition associated with an asset

Through the consumer's dashboard in the catalog browser section, this asset is offered through a Contract Offer, along with all offers within the data space. The Contract Offer provides the option to the consumer to negotiate the asset named "test-document_company1", provided by the first participant. After a short wait, a message appears validating the Contract Creation process. Thus, the data transfer process can be achieved now by clicking on the Transfer button and choosing the destination in which the actual file will be stored. The button "Start Transfer" initiates the transfer process of the actual document in the selected storage service, as illustrated in Figure 18. In Figure 19, we show the document which has successfully been transferred from the provider's storage account to the consumer.

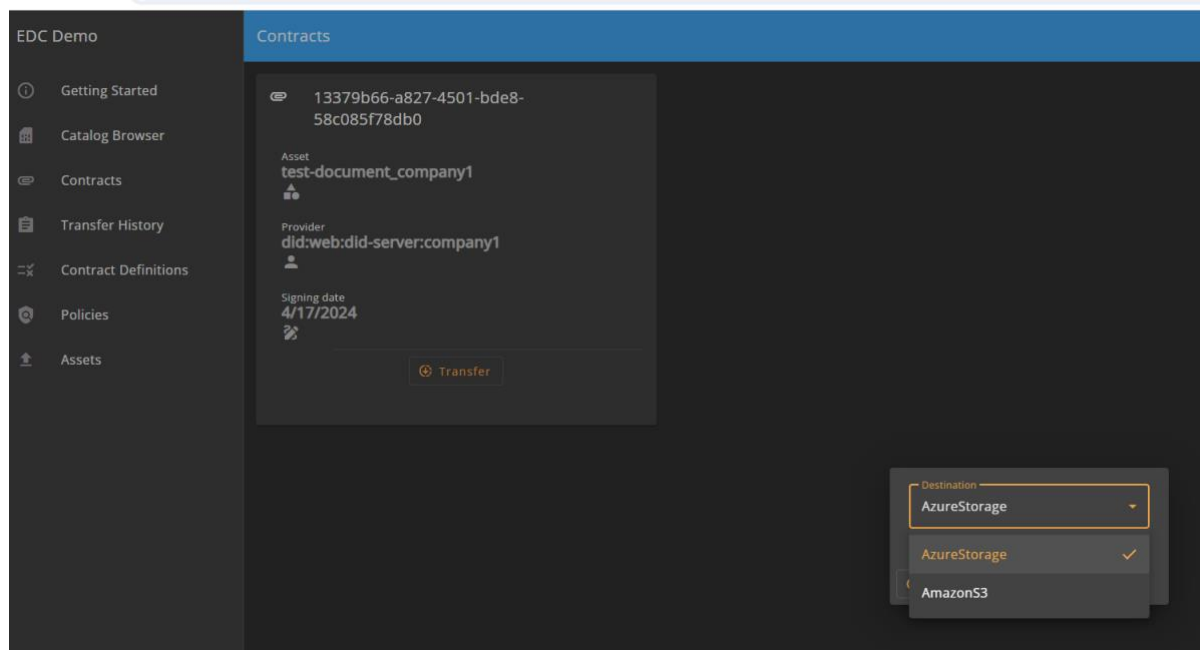


Figure 18: Transfer process

This example highlights the core functionalities of the Eclipse MVD: (i) participants can discover and access assets offered by other participants, (ii) contracts are automatically negotiated and established for data transfers, and (iii) secure data transfer is facilitated between participants.

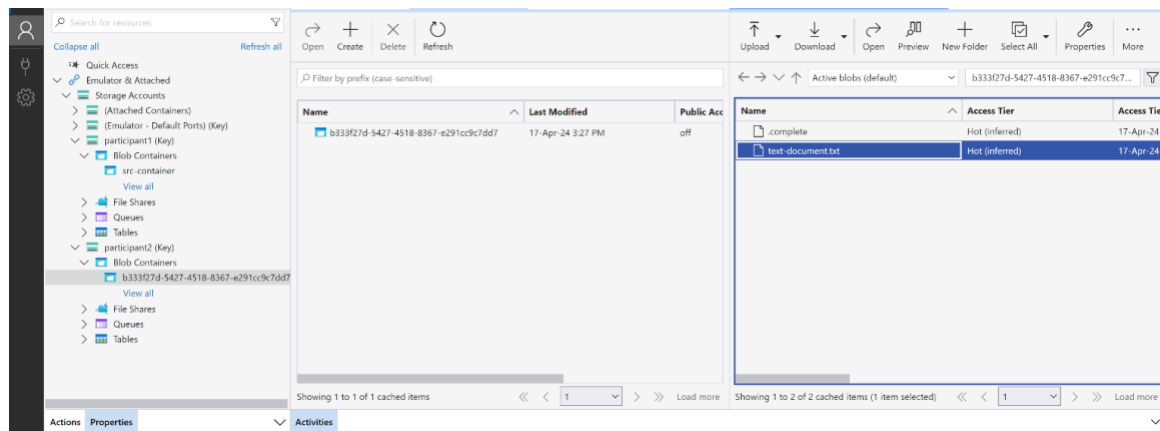


Figure 19: Consumer's storage account

3.6 Comparison of connector deployments

Based on the exploration of the different deployment implementations, a comparative table of the existing features and functionalities offered by each deployment has been created (Table 2). The comparison is based on the different possible configuration setups that were tested, the type of data that the provider can offer, the supported ways for file exchange, the existence or not of identity management functionalities, and the available Policies.

In terms of deployment, the distributed paradigm is needed for the NIKH platform, and, as far as supported files and file exchange ways are concerned, the NIKH users should be able to offer and exchange actual data files, possibly via an object storage. With regards to identity management, the support of those functionalities is crucial for the authentication and authorization of connectors participating in the NIKH data space, and lastly, the policies that are required to be supported within the NIKH, based on the defined privacy levels of the metadata records, are the Public, Public after embargo, Restricted and Closed policies. In Table 2, the support or absence of these features per connector implementation can be viewed.

Table 2: Comparison of connectors deployment

	Core Connector	Deployment	Files supported for exchange	Data Storage	Authority and Additional Services	Policies	Federated Catalogue
Single-Connectors MVD	Eclipse Data space Connector	Centralized Non-containerized	REST response with a JSON object	Local file system	No authority services	Public	NO
IDS MVD	IDS Connector	Centralized Containerized	Actual dataset exchange	Local file system	No authority services	Public	NO
Sovity MVD	Eclipse Data space Connector	Centralized Containerized	REST response with a JSON object	Local file system	No authority services	Public Public after embargo Restricted Closed	NO
Eclipse MVD	Eclipse Data space Connector	Centralized Containerized	Actual dataset exchange	Object storage service (Azurite)	Registration Service Identity Hub DID Server	Public Closed	YES

4 NextGEM network provisioning platform

4.1 Integration of data space principles in the NextGEM architecture

During the writing of the proposal, as reflected in the DoA, Data space technology was intended to provide a trusted environment for the operations in experimental studies where data could be exchanged among partners of the business network. To this end, the role of Data space is to enable a mechanism for (i) defining the access rules for data-by-data providers and (ii) facilitating the actual access to the data-by-data consumers. The envisioned data governance mechanism in the NextGEM platform supported by Data space technology offers a trust platform for enabling secure data exchange between data providers and data consumers of the platform, where the data remains at the data owner's premises and exchanged using peer-to-peer communication mechanisms, and the data access rights are declared and verified, using mechanisms governed by data owners. Following the IDS architecture, data owners remain in full control of their data and might change the terms of their usage at any time.

4.2 NextGEM data space topology

The core purpose of a Data space topology is to enable controlled, sovereign, and secure exchange and sharing of data between stakeholders. For this purpose, the IDSA introduces a decentralized data sharing architecture framework, in which data physically remain at their source and are only transferred to another participant when data exchange requests are instantiated. In this framework, data sovereignty and trust are established since each participant can attach usage restrictions to their data and monitor data transactions through continuous monitoring. Additionally, security is ensured through the identity evaluation of each participant. Additionally, this framework offers metadata storage, as well as metadata query functionalities that enable participants to search for the appropriate data sources and request access to the respective data. To this end, the NextGEM platform aims to encompass the architectural principles of trust, data sovereignty, transparency, interoperability, and integrability of diverse data sources and services.

4.2.1 Data space premises

Research organizations and institutes that are part of the NextGEM consortium maintain their own procedures and policies for the secure data storage and exchange of their research outputs. Those organizations possess different levels of available infrastructure, and consequently, have different needs regarding physical resources. It is up to the organization's administration to decide what infrastructure and how the required services are deployed to support their needs. Accordingly, the proposed topology supports the following approaches for the configuration and deployment of the physical infrastructures.

- **Local premises.** This approach consists of all necessary entities deployed on the organization's infrastructure and connected to a communication network, facilitating data exchange between data providers and consumers. To this direction, the local premises infrastructure guarantees the protection of sensitive data, as well as facilitates seamless data transfer and sharing between all involved partners. By complying with this approach, all stakeholders ensure that their data remains secure and protected, while also fostering efficient collaboration and sharing practices amongst all involved parties.
- **Limited local premises.** This approach consists of all necessary entities deployed on a centralized infrastructure and connected to a communication network, facilitating the transfer of experimental data from a data producer to other participants through a common infrastructure. In this direction, the limited local premises uses a common pool of resources that enable the deployment of isolated environments for supporting the deployment of required services for collecting and transferring data. By complying with this approach, all stakeholders can ensure that their data remains secure and protected by a service provider, while also fostering continuous interaction with other organizations.

4.2.2 Data space connector

Standardized data exchange between participants is the fundamental aspect of a data space ecosystem. The Data space Connector is the core functional component for this purpose. The IDS reference architecture model offers a flexible deployment strategy for each involved connector. A concrete deployment of a connector may differ from this structure, as existing components can be modified, and additional components may be added.

In the context of the NextGEM platform, two main approaches for supporting data exchange have been taken into consideration for supporting the deployment of a data space connector. Connectors can be further distinguished into on-premises connectors and off-premises connectors.

- **On-premises connector.** In this approach, the data space connector is deployed on the organization's premise. An on-premises connector executes the exchange of data between participants of the data space ecosystem, enabling the interaction with all connectors that belong to other organizations. Each on-premises connector provides data through the APIs it exposes. Applying this principle, there is the need to deploy a separate data storage component of each connector facilitating storage capabilities for each organization.
- **Off-premises connector.** In this approach, the data space connector typically operates in a centralized infrastructure that supports the deployments of services related to the data space connectors.

4.2.3 Data space data storage

The proposed architecture does not require centralized data storage capabilities. Instead, it pursues the idea of decentralization of data storage, which means that data physically remains with the respective data owner until it is transferred to a trusted party. This approach requires a holistic description of the data source and data as an asset combined with the ability to integrate domain-specific vocabularies for data. To manage and store the unstructured data generated from experiments, we require an object storage technology. Object storage can handle vast amounts of data in a scalable manner, allowing for efficient retrieval of unstructured data. To this end, we have integrated Azure storage into our architecture to manage this workload.

4.2.4 Data space network

The inclusion of a network in the NextGEM platform for data governance purposes aims to provide a comprehensive framework in which data ownership and data distribution policies become an integral part of the entire platform. It provides various governance and aims to facilitate secure exchange of information between participants in the NextGEM platform. Any NextGEM member can access the data governance components and publish a dataset under certain consumption rules through a network. In the same way, organizations that want to take advantage of a dataset published in the NextGEM network can request access to it and rely on data governance policies to be validated when necessary.

4.3 Data space orchestration

4.3.1 Controller

The Controller of the NIKH Platform is a component responsible for orchestrating and managing received requests. The Controller contains all of the platform's interfaces, through which a user interacts with the platform, offering the REST API endpoints. Every request is handled by the Controller, which performs the corresponding actions and generates the appropriate response.

Regarding its architecture, Controller implements various software components, which are responsible for managing the various heterogeneous system resources that are available through other services on the platform. The main functionalities of the Controller are summarized below.

- The Controller provides access to the available resources, based on the authorization/authentication token of the user (operated by the Authentication subcomponent).
- When a non-certified user wants to access available data from a NextGEM member, the Controller handles this request by initiating the communication between the NIKH Platform Connector and the respective data provider's Connector. Additionally, the Controller provides access to the open-access data of the NIKH Platform or third-party repositories (operated by the Authentication subcomponent).
- The Controller provides the available data catalog (by storing the respective metadata in its database) to the NIKH UI (Metadata subcomponent).
- The Controller provides functionalities for performing operations on the records comprising the metadata catalog (specifically for the creation, editing, and deletion of records- Metadata subcomponent)
- When a user wants to upload a new data resource to the NIKH Platform, the Controller can automate the procedure by making the appropriate API calls to the user's Data space Connector (operated by the Connector sub-component).
- The Controller can additionally ensure that the data resource is registered, as an available resource, at the Broker (Connector subcomponent).
- Finally, the Controller enables access to existing knowledge bases such as Zenodo and other 3rd sources to provide an integration to the latest public information and results (Connector subcomponent).

4.3.2 Data space orchestration

Additionally, the NIKH Controller implements a web service called Connector Manager that provides access to the data space by acting as a wrapper for the most crucial functionalities offered by the data space connectors. This enables connector interfaces to be hidden from the user, as there is no direct access. Furthermore, any information regarding the connectors is not known to the end-user or other components of NIKH. Instead, the details are acquired from the NIKH connector registry by looking up the organization the user is registered to (Figure 20). The information stored for each connector includes its address and Azurite storage attributes, essential for uploading and transferring assets. So far, the Connector Manager supports the creation, retrieval, and update of policies, contracts, and assets by consuming the APIs already provided by the IDS connector service. File uploading is not yet supported by IDS connectors, so the feature is provided by the Connector Manager which directly interfaces with Azurite instances. Since the connector has to be aware of the file and its location, this functionality is incorporated with asset publishing that already exists in the connector APIs.

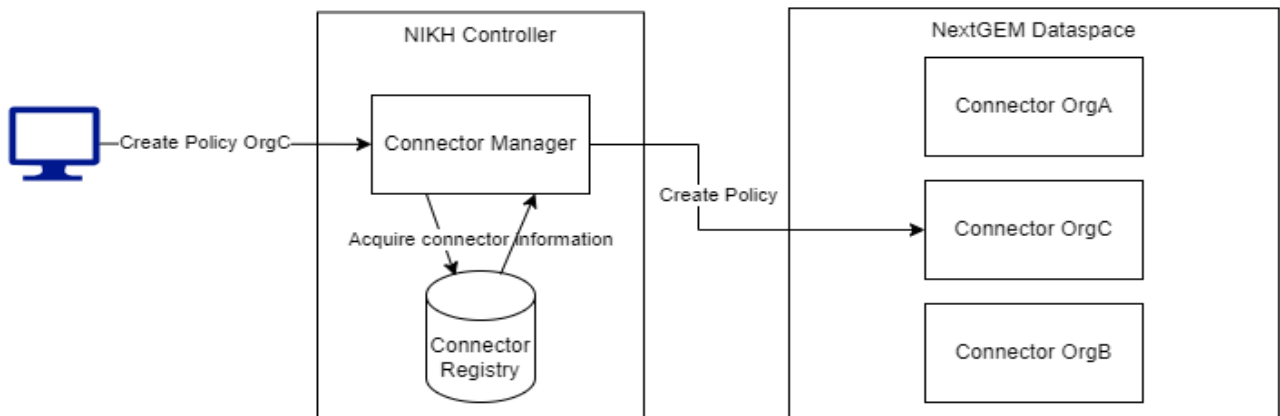


Figure 20: Connector information is acquired dynamically

The Connector Manager also facilitates a core concept of data spaces, which is file sharing between connectors. The connector deployment currently used in the NIKH platform, employs a process involving several steps, including contract negotiation and acceptance. While this procedure can be lengthy and requires calling multiple APIs, the Connector Manager encapsulates all these steps in a single function call using the API `{CONNECTOR_MANAGER_IP}/dataspace/asset/transfer`. The API is simple, requiring only the asset to transfer, the requester organization, and the provider organization. The relevant processes and the steps are depicted in the sequence diagram in Figure 21.

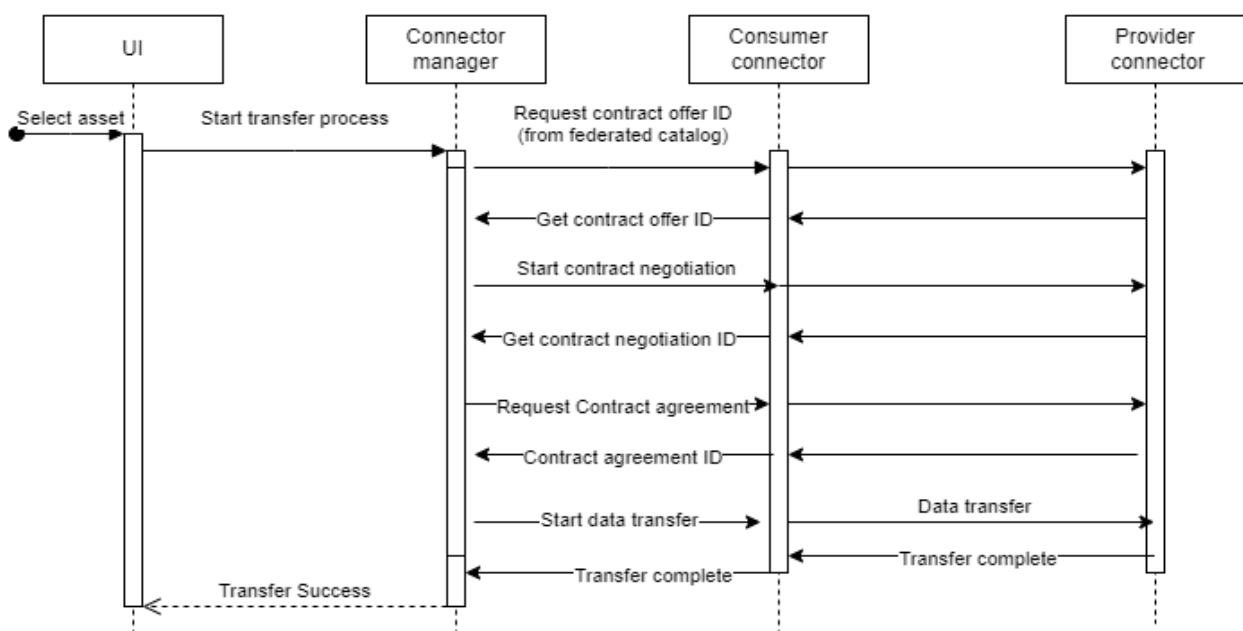


Figure 21: Asset transfer sequence

4.4 Interoperability between data spaces and services

Interoperability between Data Spaces and Services is essential for creating a connected and efficient digital ecosystem that facilitates the seamless and secure exchange of health data across various systems, platforms, and applications. Health data comes from a wide range of sources, including electronic health records, medical devices, wearables, laboratory results, public health databases, patient-reported outcomes, health insurance claims, and even experimental study results. For interoperability to function effectively, these diverse types of data must be accessible and usable across different systems and providers, regardless of their origin. This exchange of data improves decision-making, continuity of care, and collaboration across health researchers and providers. For instance, integrating data from clinical trials and experimental studies with real-world health data allows healthcare providers to assess the efficacy and safety of treatments, leading to more informed, evidence-based decisions.

The integration of health data and results from experimental studies is crucial for advancing medical research, particularly in the field of precision medicine. By enabling the secure sharing of diverse datasets, including clinical trial results, experimental study findings, genomic data, and treatment outcomes, interoperability helps researchers gain a deeper understanding of diseases and develop more targeted therapies. It accelerates innovation by connecting data from various sources, allowing for new approaches to diagnosis, treatment, and disease prevention. Moreover, interoperability improves healthcare efficiency by reducing the need for duplicate tests, minimizing administrative burdens, and preventing errors caused by incomplete or inaccessible information. As advanced technologies like artificial intelligence and machine learning become more integrated into healthcare, interoperability plays a key role in enabling these tools to analyze vast amounts of health data, driving even greater breakthroughs in healthcare. In summary, interoperability between health data spaces and services enables the seamless integration of clinical data, experimental study results, and real-world health information, improving healthcare delivery, fostering innovation, and enhancing patient outcomes across the globe.

In order to achieve interoperability between the various data sources of the NIKH platform, the data are combined into a common database. In that way, the user is able to access the catalog of available data, which includes data from all data sources. When a user uploads metadata from Zenodo, the mandatory metadata fields can automatically be retrieved by NIKH and the user can expand them with additional metadata. A metadata record is then created in the common database. When a user uploads data files, along with metadata, the metadata are also stored in the same database and the uploaded files are handled by the NIKH Connector Manager, in order to store them at the user's local premises and create the respective Asset in the user's Connector. These data can then be exchanged with other participants through the provider's and consumer's Connectors, in the sovereign and trustworthy manner that the Data space Services offer. In that way, a user can browse through the catalogue of all available metadata records, irrespective of where these metadata were retrieved from (third parties e.g., Zenodo or directly uploaded by a user along with data files).

5 NextGEM security and privacy

5.1 Data access controls and trustworthiness

5.1.1 FAIR

As stated in the DMP-interim version (D1.6: Data management plan, interim version), “the management of data and results are based on practices and procedures that ensure that used data and results are: a) stored securely and preserved in order to ensure its continuing utility, b) appropriately identifiable, retrievable, and available when needed, and c) kept in a manner that is compliant with legal obligations, including the Data Protection Act 1998 / The GDPR (Regulation (EU) 2016/679) and the Freedom of Information Act 2000 and d) able to be made available to others in line with appropriate ethical, data sharing and open access principles, especially when the data underpins published research.”

Moreover, as “data and outcomes shared in open domain can be very beneficial to society, NextGEM needs (and aims) to balance openness and protection of sensitive data carefully. As stated in the Guidelines on FAIR Data Management, data should be “as open as possible and as close as necessary”. All partners and especially data providers that participate in the consortium should comply with all applicable data protection or similar laws regulating the processing of any personal data.” The data will either be raw or processed; the latter is particularly important in the case of outputs related to personal data; concerned datasets will be anonymized.

As an EU funded project, NextGEM will participate in the Open Research Data Pilot of the European Commission (openAIRE), which enables open access and reuse of research data. NextGEM will follow an openness approach regarding its generated/gathered data and research outputs.

However, the project will employ the policy of providing data as open as possible and as closed as necessary following rules, regulations and suggestions that protect the provision of data to make results replicable and in turn protect the interests of the members of the NextGEM consortium.

5.1.2 Policies

The Eclipse Data space Connector allows users to create metadata records of their data and link them to specific usage policies through contract definitions. The access policies supported by the NextGEM Members connectors are mapped to the defined policies within NIKH, such as

- **Public.** Actual data is fully accessible to other participants with no restrictions.
- **Public after embargo.** Actual data is publicly available after a specified embargo period, while metadata remains fully accessible to other participants.
- **Restricted access.** Actual data is restricted and may require negotiation, but metadata is publicly available.
- **Closed.** Actual data and metadata are completely closed to other participants.

The respective contract definitions can be negotiated with other data space members who request access to these data. If the negotiation is accepted by both parties, a contract is agreed upon, and the data asset can be securely exchanged.

5.1.3 Access control

NextGEM has deployed and is maintaining the Keycloak ⁵⁹ authentication/authorization mechanisms for the NIKH platform. Communication of the Keycloak component is realized by providing RESTful APIs to expose its services with JavaScript Object Notation (JSON) ⁶⁰ mainly being used for the format of the data.

Keycloak is an open-source tool, licensed under Apache License 2.0, used for identity and access management. Keycloak supports a Single-Sign-On mechanism for centralized user management, while also providing feature for user roles, groups and permissions customization. Within the context of NIKH, Keycloak serves as the central authentication and authorization service, allowing for secured communication and interaction with NIKH's services.

⁵⁹ <https://www.keycloak.org/>

⁶⁰ <https://www.json.org/json-en.html>

For the purposes of the NIKH platform, we have created a new realm, named “NextGEM”, which encapsulates all the users, groups, and roles that are part of the NIKH platform. A realm is a term coined within Keycloak to represent a group of individuals or organizations that operate within a shared context and have similar interests.

Keycloak permits the creation of groups within a realm to encapsulate users under a single entity for easier management. It provides the ability to define roles in a group, which are then automatically attached to a user assigned to that group. Roles can also be created on the realm level, but assigning roles to a user in the realm has to be manual, via Keycloak’s GUI. For the needs of NIKH we have chosen the latter approach and defined three role levels: team member, full member, and supervisor (as seen in Table 3). The role of the supervisor is manually assigned by the master user of Keycloak while the other two are selected by the users upon registration and assigned using the Keycloak API when the registration request is accepted.

Some of the data that are handled by NIKH can be sensitive and should not be available for public or unauthorized access. For that reason, we have introduced the registration request in the platform. The registration request works like a normal registration process, but the user input is temporarily stored in a database until the corresponding administrator accepts or rejects the request. If the administrator accepts the request, then a new user account is created on Keycloak, and the user can then log in to the NIKH platform using the credentials defined in the registration request. If the administrator rejects the request, no further action is taken.

Table 3: User roles and permissions

	Create/Upload	Search	View	Edit	Delete
Supervisor	✓	✓	✓	✓	✓
Full Member	✓	✓	✓	✓	
Team Member		✓	✓		

The user interface does not directly communicate with Keycloak. Instead, it interfaces with a web service API (*Controller-Auth*) that encapsulates the business logic needed for user authentication/authorization and registration. The business logic contains the functions that ultimately request and receive data from Keycloak in JSON, which in turn feeds back to the user interface after some processing.

5.2 Security concerns and countermeasures in NextGEM platform

Our focus regarding security is centered on safeguarding the following:

GDPR Requirements. The GDPR represents a cornerstone in data protection legislation, enforced by the European Union (EU) to safeguard the personal data of individuals within the EU and European Economic Area (EEA). Implemented on May 25, 2018, replacing the Data Protection Directive 95/46/EC, GDPR governs principles including:

- **Lawfulness, fairness, and transparency:** Ensuring organizations have legitimate reasons for collecting and using personal data, with transparent communication and obtaining clear consent from individuals.
- **Purpose limitation:** Collecting and utilizing data solely for communicated, specific purposes.
- **Data minimization:** Utilizing only the minimum necessary personal data for specified purposes.
- **Accuracy:** Maintaining accurate and up-to-date personal data.
- **Storage limitation:** Establishing clear guidelines for data retention and deletion, avoiding unnecessary storage.
- **Integrity and confidentiality:** Implementing appropriate technical and organizational measures to protect personal data from unauthorized access, disclosure, alteration, or destruction.
- **Accountability:** Organizations are accountable for GDPR compliance, potentially involving appointing a DPO and conducting Data Protection Impact Assessments (DPIAs) for high-risk processing activities.

CIA Principles. The CIA Triad forms a foundational framework in information security, comprising three core principles to safeguard data assets:

- **Confidentiality:** Ensuring information is accessible only to authorized individuals through encryption, access controls, and data classification.

- **Integrity:** Maintaining data accuracy and reliability, preventing unauthorized or unintended modifications through techniques such as checksums, digital signatures, and access controls.
- **Availability:** Ensuring data and resources are accessible to authorized users, minimizing disruptions through redundancy, disaster recovery planning, and robust network infrastructure.

The CIA Triad provides a comprehensive framework for evaluating and implementing security controls, protecting information assets from threats like unauthorized access, data breaches, and service interruptions, guiding organizations in developing and maintaining effective information security practices.

5.2.1 Component security

Software Vulnerabilities: To effectively identify potential software vulnerabilities, it is imperative to maintain continuous monitoring for such weaknesses and promptly address them through rigorous patch management protocols. Implementing robust secure coding practices and conducting regular security audits are vital measures to mitigate the risk of exploitation. Notably, the Security and Privacy Assurance (SPA) tool, as outlined in D6.3: Trustworthy data management and compliance with ethics and legal aspects - Initial report, encompasses a suite of capabilities, including vulnerability assessments against the national vulnerability database. This feature enables the proactive identification of potentially vulnerable software. By leveraging this tool, we can promptly identify and assess any discovered vulnerabilities, allowing for timely mitigation actions to prevent potential exploitation and associated risks.

Component Communication: Regarding component communication, leakage of sensitive data to unauthorized parties is a serious concern. For this reason, secure communication protocols, which inherently use data encryption at rest and in transit, are used to protect data transmitted between components so that unauthorized data access and interception are prevented during communication.

5.2.2 Network security

Authentication and Authorization: Authentication and authorization are critical aspects for all NIKH components, pivotal for maintaining the confidentiality, integrity, and availability (CIA) of our data, as well as ensuring compliance with GDPR. Weak authentication mechanisms pose a significant risk to our data security and regulatory compliance efforts. To mitigate this risk, we implement robust authentication methods, including private-public key authentication and stringent password protection measures. These mechanisms bolster our security posture by fortifying access controls and safeguarding against unauthorized access to sensitive information.

In distributed systems such as the NIKH ecosystem, ensuring robust network security is paramount to safeguarding secure communication and data exchange among nodes. Here are the key concerns and corresponding countermeasures:

Man-in-the-Middle Attacks: Malicious actors exploiting vulnerabilities could intercept confidential information. To counter this threat, we employ strong encryption algorithms and digital certificates to authenticate and encrypt communication channels between NIKH's distributed components. Additionally, secure key management practices are implemented to prevent unauthorized interception and tampering.

Real-time Intrusion Detection: It is also a powerful characteristic of the SPA tool, which allows us to identify SSH logins originating from IP addresses not whitelisted in advance. This functionality enables us to swiftly detect and respond to suspicious activities or anomalies within the NIKH network, bolstering our security posture.

Denial of Service (DoS) Protection: Implementing effective DoS protection measures is critical to mitigate the impact of attacks targeting distributed nodes. Strategies such as rate limiting and traffic shaping mechanisms help in managing and mitigating DoS attacks. Additionally, leveraging distributed denial of service (DDoS) protection services and traffic scrubbing centers assists in filtering out malicious traffic before it reaches critical components, enhancing the resilience of the NIKH ecosystem against such threats.

Availability Monitoring: Continuous availability monitoring of the NIKH components is essential to ensure uninterrupted service delivery. By leveraging the SPA tool solution, we can identify, measure, and address potential system availability issues.

5.2.3 SPA Tool integration with NIKH components

Within the scope of NextGEM, two kinds of event captors will be enabled on the VMs that house NIKH components so that they can be adequately monitored: The Availability Event Captor and the Internet Protocol (IP) Confidentiality Event Captor.

Availability Event Captor. This captor is based on pinging or curling the respective asset (by IP or URL) to observe if the monitored service is up or down. It ensures continuous availability monitoring and helps in promptly identifying any service disruptions.

IP Confidentiality Captor. This captor gathers relevant information regarding Secure Shell (SSH) logins that occur on the targeted component. By whitelisting a set of allowed IP addresses, notifications can be provided when an unlisted IP address attempts to connect to the targeted asset through SSH, thereby enhancing security by preventing unauthorized access attempts.

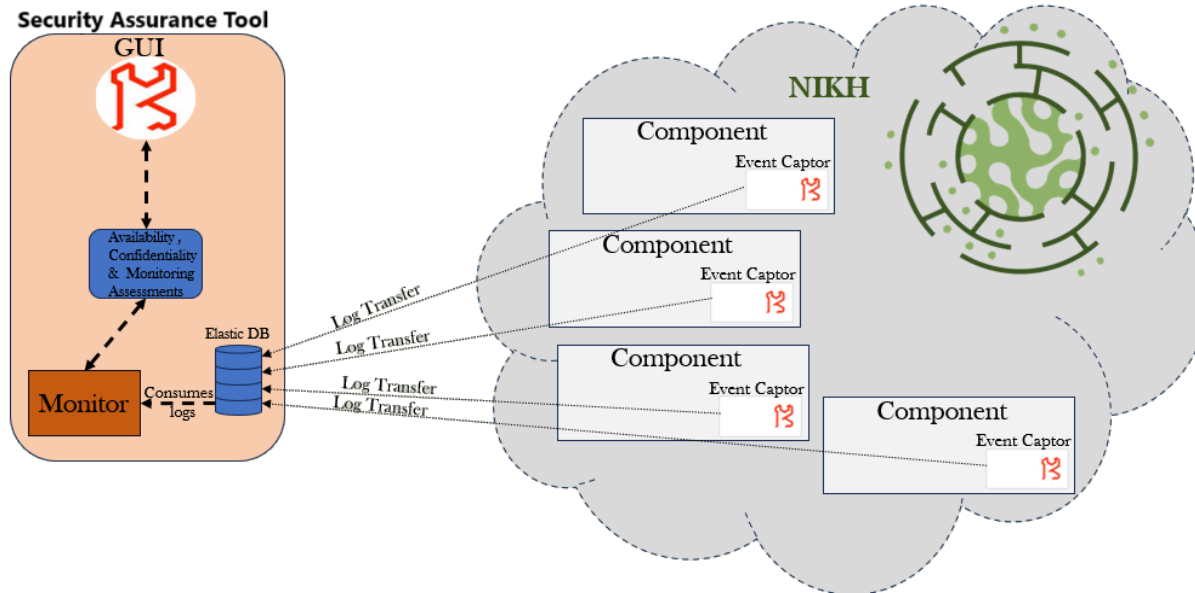


Figure 22: Shipping NIKH logs to the SPA tool

As shown in Figure 22 the aforementioned captors, ship the captured information to an Elastic Database which can be accessed and consumed by the SPA tool's monitoring component. This component is responsible for notifying SPA tool users of possible violations on the monitored components. This architecture ensures that any potential security incidents are swiftly identified and addressed, maintaining the integrity, confidentiality, and availability of the NIKH components.

5.2.4 SPA Tool vulnerability assessments and monitoring

The SPA tool enhances the NIKH platform's security through targeted vulnerability assessments and monitoring capabilities. Building on the foundational event captors described in Section 5.2.3, this section details how the SPA tool leverages these inputs for proactive risk management, supplemented by a planned penetration test to further validate the platform's defenses. These efforts ensure compliance with GDPR and uphold the CIA triad principles of confidentiality, integrity, and availability. The figures provided in this section serve as illustrative examples, as the SPA tool is undergoing continuous development. The final graphical user interface (GUI) may differ from what is currently depicted, with the latest figures to be presented in D6.8: Trustworthy Data Management and Compliance with Ethics and Legal Aspects - Final Report.

Vulnerability Assessments: The SPA tool integrates with the National Vulnerability Database (NVD) to conduct detailed vulnerability assessments of software assets defined within the NIKH ecosystem. As shown in Figure 23, these assessments produce a comprehensive report listing asset names, the affected CIA property (Confidentiality, Integrity, Availability), and a normalized likelihood score indicating the probability of exploitation. This scoring system allows the NextGEM team to prioritize mitigation efforts based on risk severity. For instance, Figure 24 highlights an NVD assessment finding for a specific software asset, providing a detailed description of the vulnerability, including exploitation methods (e.g., remote code execution) and potential impacts (e.g., data breaches or service disruptions). Leveraging this insight, proactive measures, such as applying security patches, upgrading to unaffected software versions, or adjusting configurations, can be implemented swiftly to eliminate vulnerabilities and strengthen the platform's resilience.

Assessment ID ↑↓ ▾	Assessment type ↑↓ ▾	Asset ID ↑↓ ▾	Asset name ▾	Property ↑↓ ▾	Normalised likelihood ↑ 1 ▾	Initial detection ↑↓ ▾	Last checked ↑↓ ▾
> 2	CVSSv2	4	mongodb	Availability	0/100.0 ⓘ	2024-02-07 13:18:58	2024-02-07 13:18:58
> 1	CVSSv2	1	postgresql	Confidentiality	0/100.0 ⓘ	2024-02-07 13:18:58	2024-02-07 13:18:58
> 100	CVSSv3.1	1	postgresql	Confidentiality	0/100.0 ⓘ	2024-02-07 13:18:58	2024-02-07 13:18:58
> 53	CVSSv2	6	tomcat	Confidentiality	0/100.0 ⓘ	2024-02-07 13:18:58	2024-02-07 13:18:58
> 3	CVSSv2	1	postgresql	Confidentiality	0/100.0 ⓘ	2024-02-07 13:18:58	2024-02-07 13:18:58
> 50	CVSSv2	5	openjdk	Integrity	26/100.0	2024-02-07 13:18:58	2024-02-07 13:18:58
> 80	CVSSv2	4	mongodb	Integrity	35/100.0	2024-02-07 13:18:58	2024-02-07 13:18:58

Figure 23: Example list of NVD assessment results

Name:	CVE-2023-32305	Cisa Vulnerability Name:	-
Source Identifier:	security-advisories@github.com	Description:	aiven-extras is a PostgreSQL extension. Versions prior to 1.1.9 contain a privilege escalation vulnerability, allowing elevation to superuser inside PostgreSQL databases that use the aiven-extras package. The vulnerability leverages missing schema qualifiers on privileged functions called by the aiven-extras extension. A low privileged user can create objects that collide with existing function names, which will then be executed instead. Exploiting this vulnerability could allow a low privileged user to acquire 'superuser' privileges, which would allow full, unrestricted access to all data and database functions. And could lead to arbitrary code execution or data access on the underlying host as the 'postgres' user. The issue has been patched as of version 1.1.9.
Created:	2024-02-01 18:00:01		
Published:	2023-05-12 19:15:08		
Last Modified:	2024-02-01 14:06:55		
Vuln Status:	Analyzed		
Evaluator Comment:	-		
Evaluator Solution:	-	Cve References:	Url: https://github.com/aiven/aiven-extras/security/advisories/GHSA-7r4w-fw4h-67gp
Evaluator Impact:	-		
Cisa Exploit Add:	-	Source:	security-advisories@github.com

Figure 24: Example NVD finding

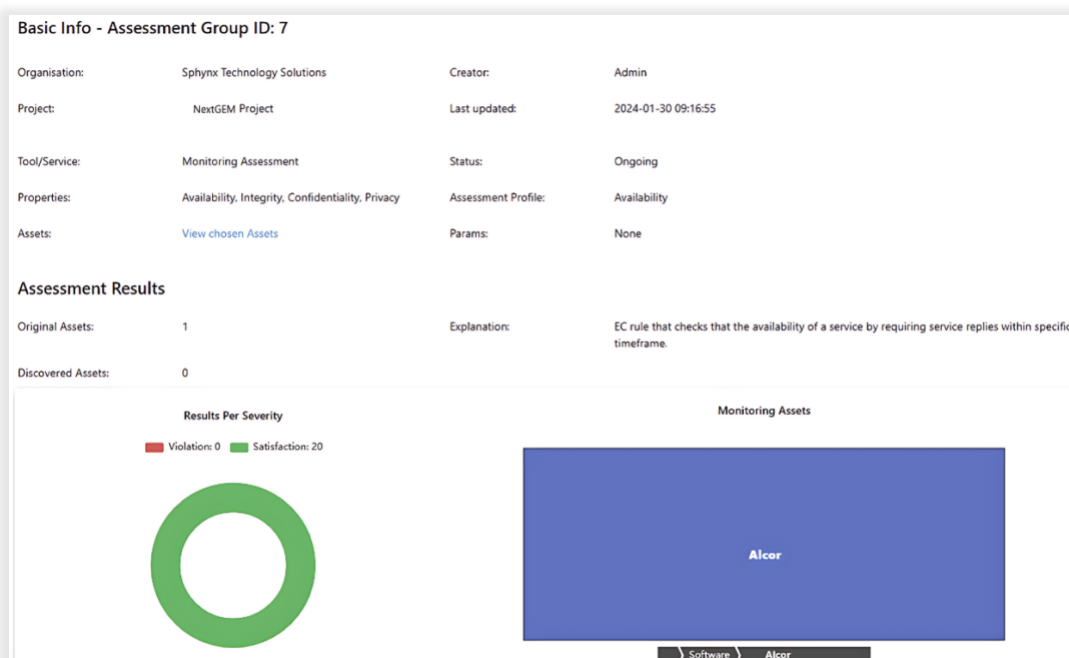


Figure 25: Availability Assessment overview

Availability Monitoring: The SPA tool performs availability monitoring by periodically pinging specific NIKH components or URLs to confirm their operational status, leveraging the Availability Event Captor described in 5.2.3. This process targets VMs hosting critical services, such as the Data space services, as well as external data spaces, to detect downtime or unresponsiveness. A predefined rule within the SPA tool is triggered if a component fails to respond, registering a violation (see Figure 25 and Figure 26). Each check's outcome, including the time of the violation or successful response, alongside relevant metadata (e.g., component identifier), is catalogued in the system. This cataloguing allows the NextGEM team to monitor availability and address incidents, ensuring the reliability of the NIKH platform's distributed infrastructure.

Assessment ID	Assessment type	Asset ID	Property	Normalised likelihood	Initial detection	Last checked	Valid until
182	Monitoring assessment	4	Availability	0	2024-01-30 09:16:54	2024-01-30 09:16:54	
<p>Assessment Criterion ID: 13</p> <p>Criterion Description: EC Availability rule that ensures an asset is available, by requiring service replies within specific timeframe (300ms).</p> <p>Result: Satisfaction</p> <p>If Event:</p> <p>Prefix: Happens</p> <p>Event:</p> <p>ID: 39</p> <p>Status: call</p> <p>Sender: EventCaptor</p> <p>Receiver: Alcor</p> <p>Source: NativeCaptor</p> <p>@Timestamp: 1706506214927</p> <p>Arguments:</p> <p>OperationName: _OPNAME</p> <p>OperationInstance: _OPINST</p> <p>arg3: _ARG1</p> <p>arg4: _ARG2</p> <p>Then Event:</p> <p>Prefix: Happens</p> <p>Event:</p> <p>ID: 40</p> <p>Status: res</p> <p>Sender: Alcor</p> <p>Receiver: EventCaptor</p> <p>Source: NativeCaptor</p> <p>@Timestamp: 1706506214927</p> <p>Arguments:</p> <p>OperationName: _OPNAME</p> <p>OperationInstance: _OPINST</p> <p>arg3: _ARG1</p> <p>arg4: _ARG2</p>							

Figure 26: Availability Assessment Result & Rule

Penetration Testing: To assess NIKH's security controls, a penetration test is scheduled as part of NextGEM's ongoing assurance activities. This test simulates real-world attack scenarios targeting the platform's distributed components, APIs, and user interfaces. The scope encompasses external threats (e.g., man-in-the-middle attacks, DoS attempts) and internal risks (e.g., privilege escalation via misconfigured roles). Unlike the automated assessments, this evaluation probes for complex vulnerabilities, such as logic flaws, that may evade detection. A detailed report will be shared exclusively with the platform's developers, outlining findings and remediation recommendations to address identified weaknesses securely. This exercise reinforces NextGEM's proactive approach to maintaining a robust security posture.

Through vulnerability assessments, availability monitoring, and penetration testing, the SPA tool delivers a multi-layered strategy for protecting the NIKH platform. These capabilities enable rapid risk identification and resolution, maintaining a secure and reliable environment for NextGEM's stakeholders. Dedicated deliverables addressing security assurance, namely D6.3: Trustworthy Data Management and Compliance with Ethics and Legal Aspects - Initial Report and D6.8: Trustworthy Data Management and Compliance with Ethics and Legal Aspects - Final Report, provide further details on these efforts. D6.3 includes initial insights into the SPA tool's capabilities, as referenced earlier, while D6.8 will present the final outcomes of the project's security assurance activities. For security reasons, specific vulnerabilities identified during assessments or testing are disclosed only to NIKH developers, ensuring that sensitive information remains protected.

6 NextGEM Services

6.1 Towards NextGEM data space implementation

The NIKH platform, as thoroughly described in the aforementioned sections and previously submitted deliverables, is considered the major appliance of the NextGEM data space implementation based on the Eclipse Data space Connector. More specifically, the NIKH platform comprises two main framework components, the *Data Storage* and the *Data Connector*, both necessary for the expected NextGEM Data space to be deployed and integrated among partners (Figure 27).



Figure 27: Premises – Connector – Data space, the three key framework components of the NIKH platform

Connector: The Eclipse Dataspace connector acts as a middleware layer that enables a seamless integration and interaction between the environments of the Data Storage and the data space ecosystem, offering enhanced data management capabilities. It allows participants to efficiently search for published data offerings of other participants via metadata and filter options. Data offerings of the participants can automatically be visible through the platform after their connector has been registered in the data space.

At its core, the connector consists of several essential sub-components and services. Below, a breakdown of the key roles and functions of a data space connector is presented:

- Data Usage Policies
- Data Contract
- Data Asset
- Data Catalogue
- Identity Manager

Data Storage: It typically refers to the underlying infrastructure used to store and manage the data assets within the platform. The ultimate goal is to keep the data with or under the control by the data owner. The choice of Data Storage can vary depending on factors such as performance requirements, scalability, data types, and compliance considerations. However, common data storage solutions used in data spaces may include:

- Object Storage
- Distributed File Systems
- (Non)-Relational Databases
- Data Lakes

For the NIKH platform, the Azurite emulator has been used to mirror the functionalities of the Azure Blob storage service within local development environments. This emulation is achieved through a comprehensive set of features, replicating the behavior of the Azure Blob storage, supporting various blob types including block blobs and page blobs, along with functionalities such as metadata handling and blob snapshots.

However, decoupling the connector and data storage, these two key components within a data space architecture (Figure 28), introduces a versatile approach that offers distinct solutions, namely *as-a-service*, as they are thoroughly described in the following sections. This separation provides organizations/stakeholders with enhanced flexibility and scalability, enabling them to tailor their data management strategy to their specific needs and requirements.

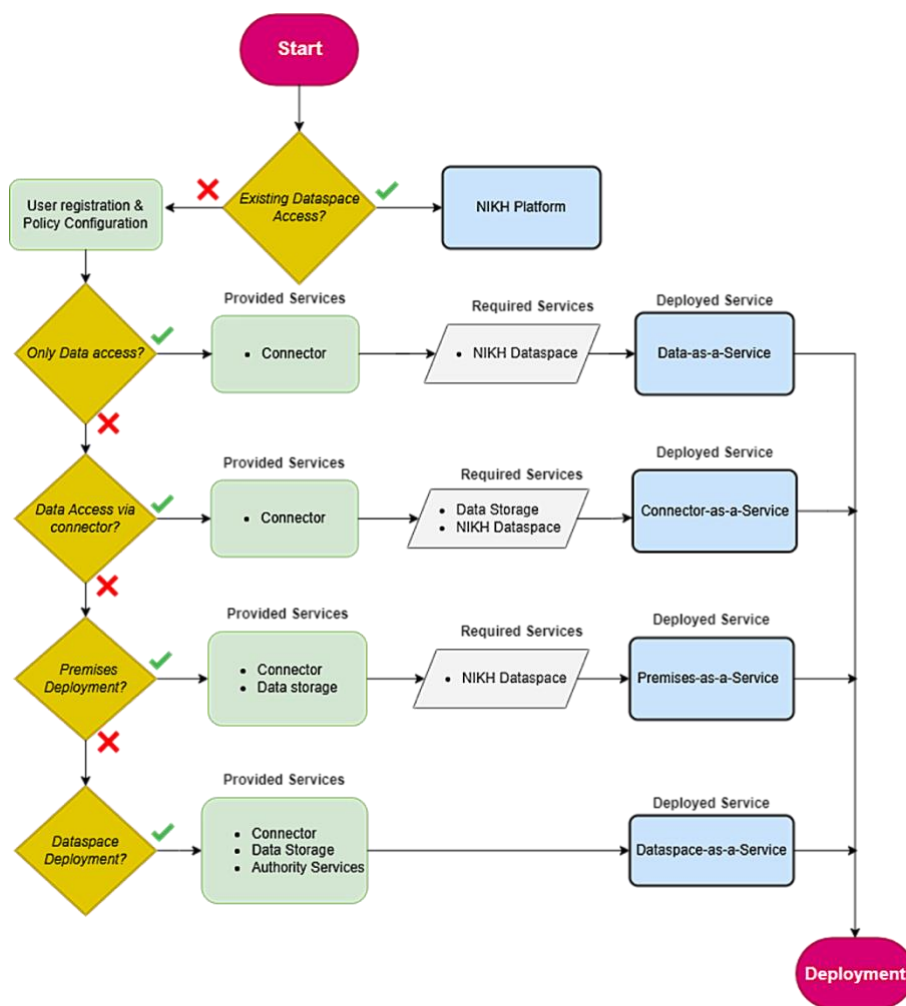


Figure 28: The relation between all 'as-a-Service' scenarios

6.2 Data-as-a-Service (DaaS)

The unified system of NIKH Platform as a service can benefit every type of potential stakeholder in getting data access to the NextGEM Data space. DaaS provides end-users and organizations with a comprehensive solution for accessing data resources within a data space ecosystem, ensuring environments of high reliability and performance. A set of web interfaces allows end-users to seamlessly interact with data space functionalities, supporting workflows for efficient data retrieval and metadata management.

Through the web interfaces, access is provided to the end-users to seamlessly interact with the data space functionality, enabling workflows for efficient data retrieval, manipulation/interaction, and metadata management. Whether querying for specific data points, browsing through extensive datasets, or collaborating with team members on shared projects, the service offers a streamlined experience tailored to the user's needs. With robust authentication, authorization, and permission controls, the service ensures secure access to their data, especially for sensitive information, to remain protected while allowing users to customize and control access to their data resources. Overall, data access as a service powered by the data space connector offers flexibility, scalability, and convenience, enabling users to unlock the full potential of their data assets for enhanced decision-making and innovation.

Ultimately, data access as a service empowers end users to harness the full potential of their data assets without the need to deploy their own data space topology and reserve any of their own resources. To enable the DaaS functionality to the end users, three major pillars are assessed and properly configured as shown in Table 4:

Table 4: Data-as-a-Service specification

Data Storage	Connector	Data space
Not needed	Provided	Required

6.3 Connector-as-a-Service (CaaS)

The data space connector plays a pivotal role in connecting and exchanging data among various parties of different entities into the data space. In brief, its primary function is to seamlessly integrate the powerful data management capabilities of the data space directly into the Eclipse environment, providing users with effortless access to a wealth of data resources. Through the connector, end users can easily discover, explore, and interact with datasets, files, and databases stored within the data space. This integration streamlines the workflow for users, eliminating the need to switch between different tools or platforms to access data. Additionally, the connector enhances collaboration between users by enabling sharing and version control of data resources within the data space. Overall, the connector empowers end-users to leverage the full potential of the data space for their data-driven tasks, facilitating productivity and efficiency in their work. As a result, with a standardized implementation of the NextGEM controller, we are capable of providing the NextGEM connector of the NIKH platform as a Connector-as-a-Service (CaaS) solution.

The primary function of CaaS is to seamlessly enable data exchange capabilities among various entities within the data space ecosystem. Using data space connectors, end-users can easily explore and retrieve heterogeneous data across the ecosystem. Consequently, CaaS provides the NIKH platform's connector to facilitate collaboration among users, allowing data sharing without the need to switch between different tools or platforms.

This solution provides the stakeholders with an easy and fast way to access the data space and an immediate way to update and provision data to the involved parties (Table 5). In addition, CaaS offers several key advantages:

Flexibility: By decoupling data storage, end users have the freedom to choose the most suitable storage solution for their needs. They can opt for a cloud-based storage service like Amazon S3, Azure Blob Storage, Google Cloud Storage, an on-premises database system, or any other storage infrastructure that meets their requirements. This flexibility allows them to leverage existing deployments in data storage infrastructure or choose the best-in-class solution for their specific use case.

Global Reach: With respect to global reach, by adopting a plug-and-play approach, CaaS might be deployed across multiple regions or data centers, ensuring global accessibility for the end users. This global reach enables flexibility in order for the connector solution to be closer to the end users, reducing latency and improving performance. Additionally, it allows candidates to reach a broader audience and seamlessly expand their operation footprint without geographic constraints.

Scalability: Decoupling data storage from the connector implementation enables end users to scale their storage infrastructure independently of the connector. They can dynamically adjust storage capacity and performance to accommodate changing data volumes and usage patterns without affecting the functionality or performance of the connector. This scalability ensures that the connector can handle growing data volumes and evolving business needs effectively.

Modularity: CaaS promotes modularity and separation of services in the overall system architecture. Each component can be developed, deployed, and maintained independently, making the system easier to understand, test, and maintain. This modular approach also facilitates future enhancements or upgrades to either the storage or connector component without disrupting the entire system.

Data Migration and Portability: CaaS approach also simplifies data migration and portability between different storage solutions. End users can easily transition from one storage provider to another or migrate data between on-premises and cloud-based storage environments without making changes to the connector.

Table 5: Connector-as-a-Service specification

Data Storage	Connector	Data space
Required	Provided	Required

6.4 Premises-as-a-Service (PaaS)

The Premises-as-a-service (PaaS) provides stakeholders and organizations with virtual premises, including reliable storage connected to a data space ecosystem solution that is both flexible and reliable solving their major problem with limited local storage resources without worrying about the complexities of deploying and managing storage infrastructure maintenance. Instead, the PaaS solution offers participants a range of storage benefits to the participants including seamless scalability, high availability, and robust security measures eliminating the need for

their resources. Participants can take advantage of avoiding using in-house resources and retaining the possibility to confidently store, manage, and access their data assets, knowing that their data is securely hosted and managed by a trusted service provider and retaining full control of them.

Although PaaS provides a scalable storage solution that adapts to changes in data volume and usage patterns to meet regulatory standards and best practices, stakeholders must join a pre-deployed, pre-configured data space environment tailored to their specific needs (Table 6).

Table 6: Premises-as-a-Service specification

Data Storage	Connector	Data space
Provided	Provided	Required

6.5 Dataspace-as-a-Service (DSaaS)

Data spaces as a whole offer a comprehensive solution to expedite the growth of the data exchange and facilitate seamless data sharing along with effective and secure data management. They present a decentralized, standardized, and trusted framework for various data-driven applications enabling smooth communication among participants and rapid scalability of different use cases by fostering interoperability. This allows for the effortless sharing of data across diverse environments, scientific or not, while providing clear guidelines on how and when data can be accessed. Sovereign data sharing within data spaces fosters innovation, stimulates the development of new digital products or services, enhances collaboration, and promotes transparency.

Leveraging all these advantages of data spaces and providing them as a service (DSaaS), it offers the flexibility to access a managed environment tailored to specific needs, determining a centralized data space authority (data space administrator) to maintain complete control over the data space. This service provides a turnkey solution for establishing and operating a data space, including infrastructure provisioning, configuration, and ongoing maintenance. With DSaaS, the setup process is streamlined and focused on the capabilities of the deployed data space to drive innovation and collaboration within the organization and with external partners.

In addition, the data space authority has the responsibility to define overarching rules, including aspects such as business models, fees, participant criteria, and purpose. Furthermore, the authority can address compliance concerns, ensuring adherence to local regulations, transparency standards, and cybersecurity protocols.

According to Table 7, for a DSaaS to be deployed, it is necessary to provided the end-users with each of the three reported components, specifying their particular operational and technical requirements.

Table 7: Data space-as-a-Service specification

Data Storage	Connector	Data space
Provided	Provided	Provided

7 Development of NIKH data space platform

7.1 Deployment overview and service topology

The NIKH data space platform establishes a comprehensive framework that enables secure and sovereign information exchange between participants through data ownership and distribution policies. To this end, various standards and protocols have been implemented to support the integration of diverse services necessary for creating a unified platform. Specifically, the proposed data space architecture includes configurations and deployments for secure data exchange within the health data space. To establish a functional data space ecosystem, the NIKH platform incorporates functionalities such as data space connectors, data storage, service orchestration, and access control, all deployed on a container orchestration system, as presented in Figure 29 and detailed below.

The NIKH platform's data space ecosystem encapsulates its main components in container images, independently deployed as Docker containers. These container images are automatically built within GitLab CI, based on configuration files (Dockerfiles and Docker Compose files) provided by component owners. The containerized NIKH software components are deployed in Kubernetes using appropriate configuration files (Kubernetes deployment YAML configurations). The Eclipse Data Space Connector, selected based on the investigation in the previous section, is built upon the Eclipse MVD and serves as the most suitable solution for representing organizations and data providers/consumers across various locations. Acting as a middleware layer, it enables seamless integration and interaction between data storage environments and the platform, offering enhanced data management capabilities. Data offerings become visible on the platform automatically once a participant's connector is registered in the data space. The ultimate goal is to keep data with or under the control of the data owner. For the NIKH platform, the Azurite emulator has been utilized to replicate Azure Blob storage service functionalities within local development environments, supporting object storage capabilities through a comprehensive feature set.

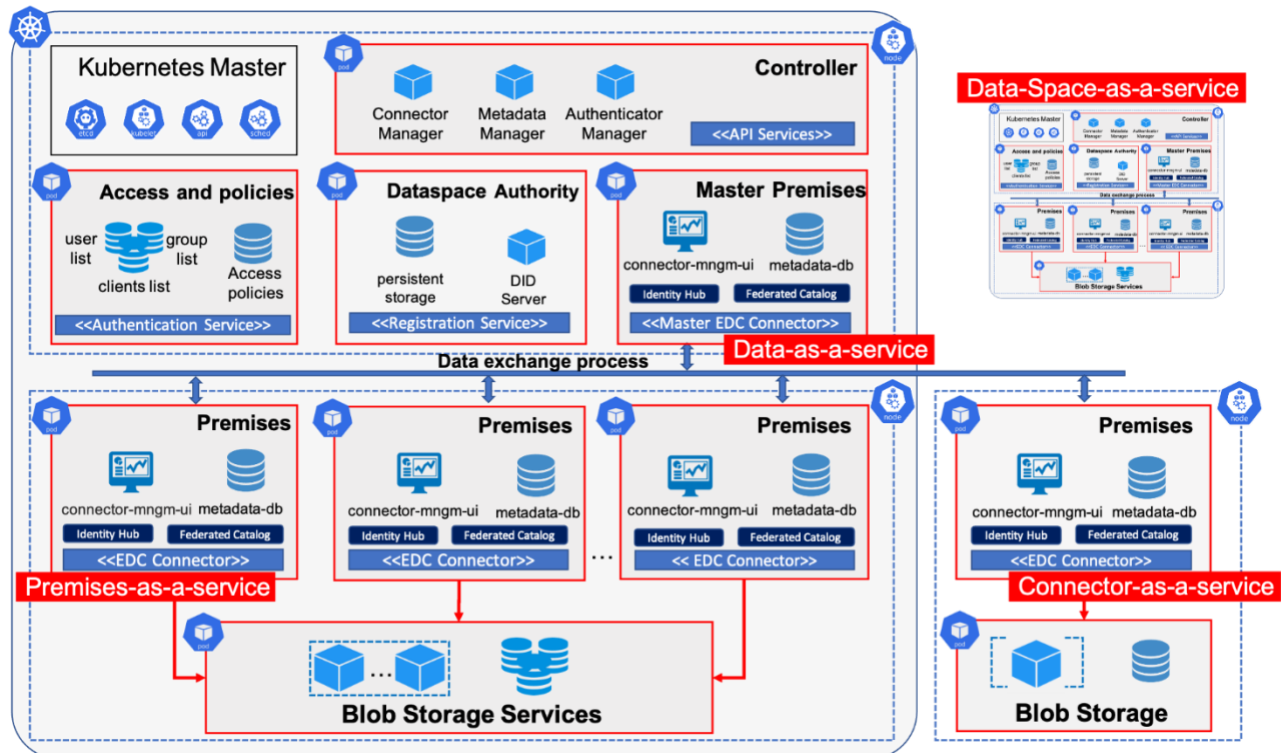


Figure 29: Deployment of NIKH data space ecosystem

To enable efficient access control policies, a harmonized mechanism was deployed for access control using Keycloak and access policies through the EDC. Specifically, the Keycloak open-source tool is used for identity and access management, providing customizable features for user roles, groups, and permissions. It has been deployed as a container on the same Kubernetes cluster, and communication with Keycloak is facilitated through RESTful APIs, primarily using JSON as the data format. To streamline user access, a registration request process has been introduced on the platform, allowing new user accounts to be created within a Keycloak realm, where users can log in to the NIKH platform. This setup enables a centralized authentication and authorization service for secure communication and interaction with NIKH's services. For access policies, the EDC allows users to

create metadata records of their data and link them to specific usage policies through contract definitions. Contract offers can be negotiated with users who request access to this data, based on the defined access policies. The available policies are: i) public, where actual data is fully accessible with no restrictions; ii) public after embargo, where data is publicly available after a specified embargo period; iii) restricted access, where data may require negotiation; and iv) closed, where both data and metadata are inaccessible to other participants.

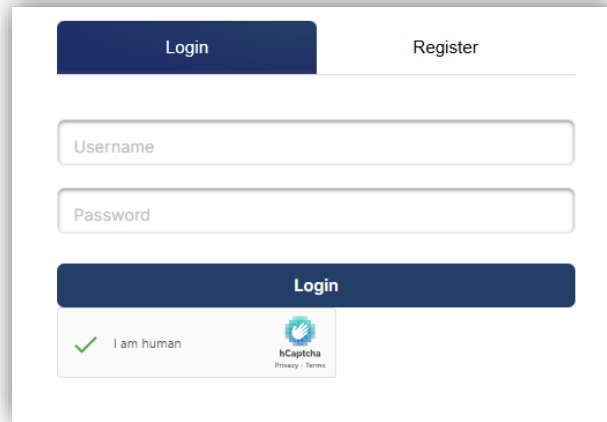
The Controller of the NIKH platform is designed to orchestrate all platform interfaces, providing REST API endpoints for the connector, metadata, and authentication manager. It integrates the deployed Connector Manager, which enables access to the data space by acting as a wrapper for key functionalities, including file sharing between data space connectors and supporting the creation and updating of policies, contracts, and assets by utilizing the APIs provided by the data space connector service. The Controller offers functionalities for authentication and manages the metadata catalog, including the creation, editing, and deletion of entries. It also automates the process of new data uploads by making the necessary API calls to the Data Space Connector. All sub-components of the Controller have been containerized and are managed within a Kubernetes container orchestration environment.

7.2 User interface

The Application Portal serves as a central access point for both internal and external NextGEM stakeholders, offering a visual interface for user interaction with NIKH and its components according to their access rights. Users engage with the portal through a GUI (see Figure 30), consisting of two main elements: a public static content page with information on government regulations, guidelines, and the current state of EMF research, and a secure, access-controlled dynamic application that interacts with NIKH services. While the GUI itself does not expose any APIs for communication with other components in the NIKH architecture, it sends requests to the Controller's APIs to facilitate interactions with various components via the relevant Managers within the Controller.



Figure 30: NIKH's Front page



The login prompt form features a dark blue header with 'Login' and 'Register' tabs. Below are input fields for 'Username' and 'Password'. A large dark blue 'Login' button is positioned below the password field. At the bottom, there is a green checkmark icon with the text 'I am human' and an hCaptcha logo with links for 'Privacy' and 'Terms'.

Figure 31: NIKH's login prompt

The NIKH dashboard, which has a robust data management system, enables users to view metadata from the NextGEM hub and/or upload new data to the platform.

The static part of the NIKH portal is offered to any user without authentication, thus providing information to the casual visitor and extending to the general public. The dynamic part of NIKH is strictly offered to the authorized subscribers as having access to NIKH's data spaces functionality. These subscribers can either be members of an organization that has premises within the NIKH ecosystem (publishers/consumers) or external members that would like to utilize NIKH's data spaces to search and obtain posted works (consumers). The validation of all subscribed NIKH members is done via username/password credential authorization (see Figure 31). A Captcha mechanism is enforced upon login, which ensures a human individual, as opposed to a bot, is attempting to access the service.

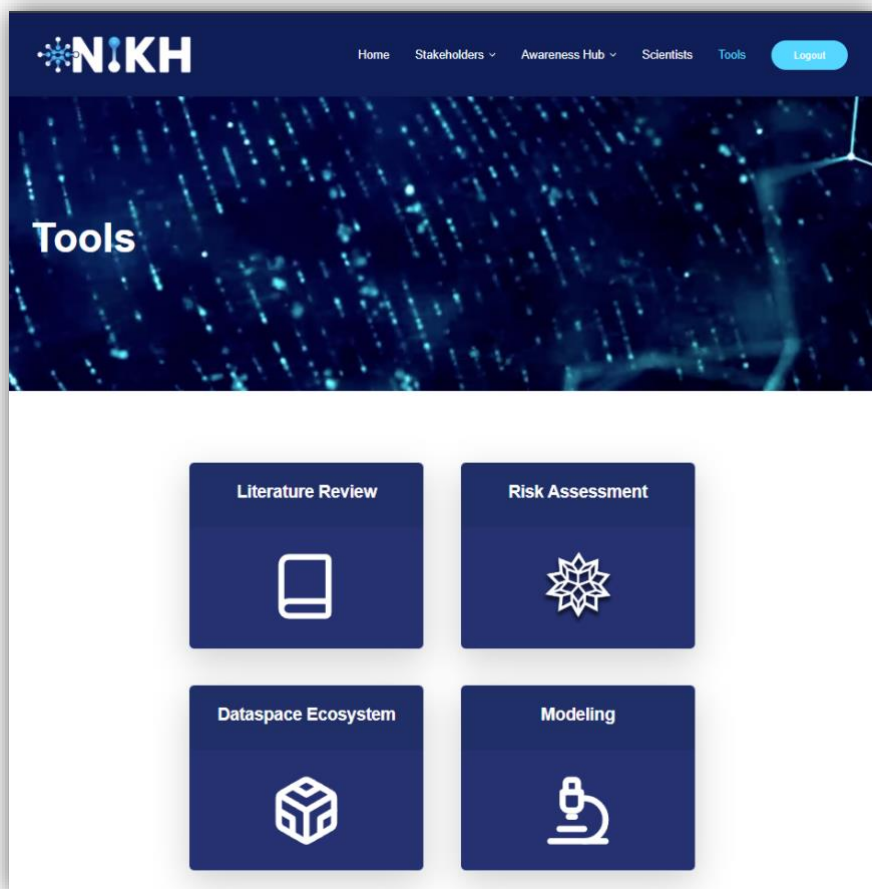


Figure 32: NIKH's Tools page

Upon login, the user is presented with NIKH's Tools page (see Figure 32), which presents four options: Literature Review, Risk Assessment, Data space Ecosystem, and Modeling. For this deliverable, we will shift our focus to the Dataspace Ecosystem option, which interfaces with the core of NIKH's data spaces.

By selecting the Dataspace Ecosystem tab, the user is presented with NIKH's Ecosystem page (see Figure 33). This page contains four tabs/options: Assets, Contracts, Premises, and Users

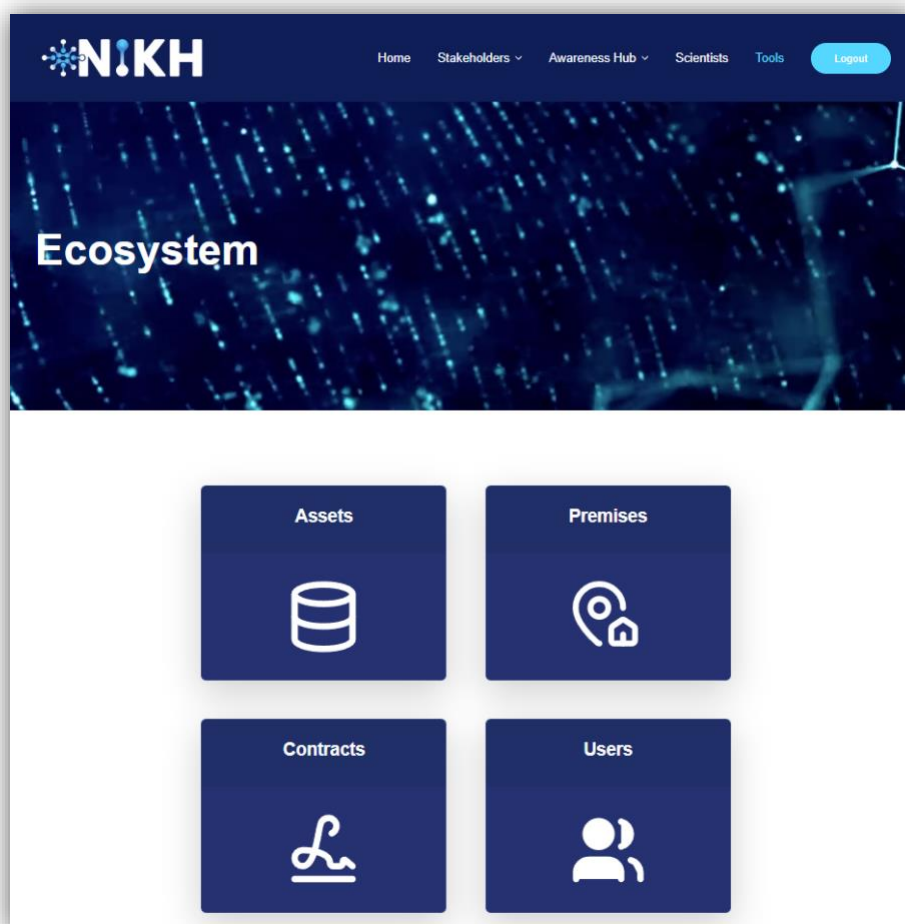


Figure 33: NIKH's Data space ecosystem page

7.2.1 Assets

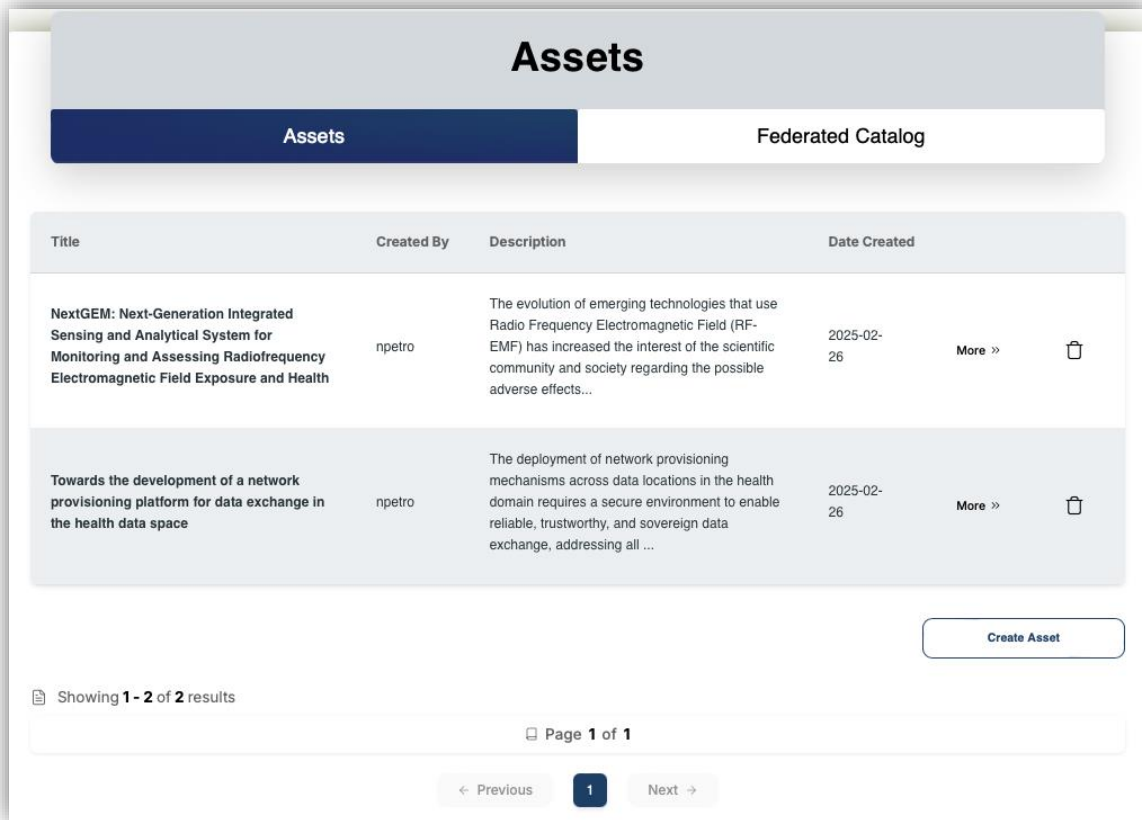
Data assets in NextGEM encompass resources shared within the ecosystem, including research papers, datasets, and results. These assets serve as the core units of exchange, enabling collaboration and innovation through secure and efficient data sharing. By contributing to and accessing these assets, members of the data space generate insights, enhance decision-making, and create value across various domains.

By selecting the Assets tab on the Dataspace Ecosystem page (see Figure 33), a visiting member can access NIKH's Assets page (see Figure 34). On this page, an ecosystem member has three options:

- View a list of owned assets
- Create a new asset
- Access the federated catalog, which includes assets provided by other premises within the ecosystem.



The Assets page contains two tabs. The left tab (see Figure 34), labeled Assets, displays the list of owned assets uploaded to NIKH from the premises that the logged-in user represents or belongs to. Within the Assets tab, users can view key details, including the title of the asset, the creator, a description, and the date of creation. Additionally, a "More >>" button provides further information about each asset (Figure 35).

For asset management, an authorized user can delete an asset by clicking on the trashcan icon at the end of each asset entry, removing it from the premises.



Assets

Assets
Federated Catalog

Title	Created By	Description	Date Created	
NextGEM: Next-Generation Integrated Sensing and Analytical System for Monitoring and Assessing Radiofrequency Electromagnetic Field Exposure and Health	npetro	The evolution of emerging technologies that use Radio Frequency Electromagnetic Field (RF-EMF) has increased the interest of the scientific community and society regarding the possible adverse effects...	2025-02-26	More » 
Towards the development of a network provisioning platform for data exchange in the health data space	npetro	The deployment of network provisioning mechanisms across data locations in the health domain requires a secure environment to enable reliable, trustworthy, and sovereign data exchange, addressing all ...	2025-02-26	More » 

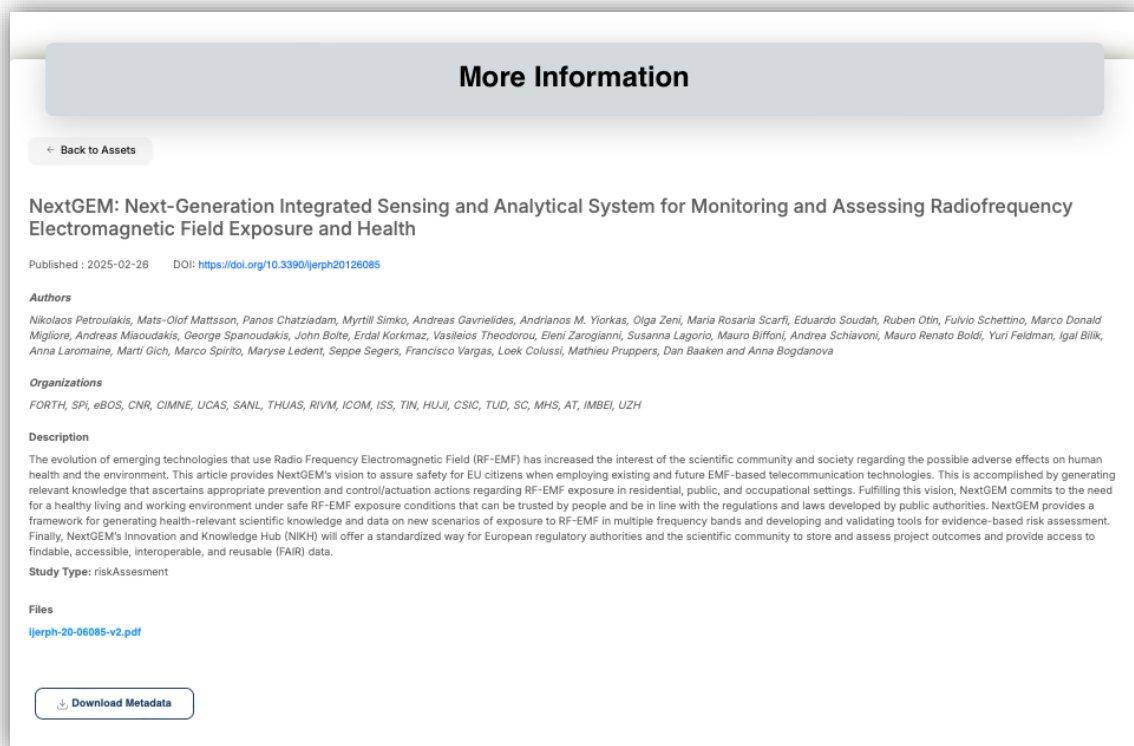
[Create Asset](#)

Showing 1 - 2 of 2 results

Page 1 of 1

← Previous
1
Next →

Figure 34: NIKH's Assets view



More Information

[← Back to Assets](#)

NextGEM: Next-Generation Integrated Sensing and Analytical System for Monitoring and Assessing Radiofrequency Electromagnetic Field Exposure and Health

Published : 2025-02-26 DOI: <https://doi.org/10.3390/ijerph20126085>

Authors

Nikolaos Petroulakis, Mats-Olof Mattsson, Panos Chatziadam, Myrtil Simko, Andreas Gavrielides, Andrianos M. Yiorkas, Olga Zeni, Maria Rosaria Scarfi, Eduardo Soudah, Ruben Otin, Fulvio Schettino, Marco Donald Migliore, Andreas Miaoudakis, George Spanoudakis, John Bolte, Erdal Korkmaz, Vasileios Theodorou, Eleni Zarogianni, Susanna Lagorio, Mauro Biffoni, Andrea Schiavoni, Mauro Renato Boldi, Yuri Feldman, Igal Blikk, Anna Laromaine, Marti Gich, Marco Spirito, Maryse Ledent, Seppe Segers, Francisco Vargas, Loek Colussi, Mathieu Pruppers, Dan Baaken and Anna Bogdanova

Organizations

FORTH, SPI, eROS, CNR, CIMNE, UCAS, SANL, THUAS, RIVM, ICOM, ISS, TIN, HUJI, CSIC, TUD, SC, MHS, AT, IMBEI, UZH

Description

The evolution of emerging technologies that use Radio Frequency Electromagnetic Field (RF-EMF) has increased the interest of the scientific community and society regarding the possible adverse effects on human health and the environment. This article provides NextGEM's vision to assure safety for EU citizens when employing existing and future EMF-based telecommunication technologies. This is accomplished by generating relevant knowledge that ascertains appropriate prevention and control/activation actions regarding RF-EMF exposure in residential, public, and occupational settings. Fulfilling this vision, NextGEM commits to the need for a healthy living and working environment under safe RF-EMF exposure conditions that can be trusted by people and be in line with the regulations and laws developed by public authorities. NextGEM provides a framework for generating health-relevant scientific knowledge and data on new scenarios of exposure to RF-EMF in multiple frequency bands and developing and validating tools for evidence-based risk assessment. Finally, NextGEM's Innovation and Knowledge Hub (NIKH) will offer a standardized way for European regulatory authorities and the scientific community to store and assess project outcomes and provide access to findable, accessible, interoperable, and reusable (FAIR) data.

Study Type: riskAssessment

Files

[ijerph-20-06085-v2.pdf](#)

[Download Metadata](#)

Figure 35: NIKH's Asset details

Apart from viewing assets, a user can create a new one using the "Create Asset" option, available as a button. Clicking the "Create Asset" button, accessible only to Supervisors and Full Members, opens a form (see Figure 36) that includes all the necessary fields for adding a new asset to the premises.

The required fields are:

- Title, the title of the study
- Authors, the authors of the study separated by commas
- Organizations, the related organization to the study, separated by commas
- Description, a longer description of what the study is about
- Free Keywords, that would help with searches, separated by commas
- The selection of Type of Study via a drop-down list, options include
 - exVivo
 - exposureAssessment
 - humanStudies
 - inVitro
 - inVivo
 - riskAssessment
 - simulation
- The Type of Output via a drop-down list, options include:
 - audio
 - codebook
 - dataset
 - deliverable
 - image
 - poster
 - presentation
 - publication
 - report
 - software
 - video
- The study actual file in any desired format, uploaded by pressing the “Upload Data” button

Aside from the above mandatory fields, the user has the option to upload additional metadata by pressing the “Upload Metadata” button. Utilizing this feature, domain-specific metadata may be uploaded, such as characteristics of the study or experiment, parameters used, etc. The type of file is up to the user’s discretion.

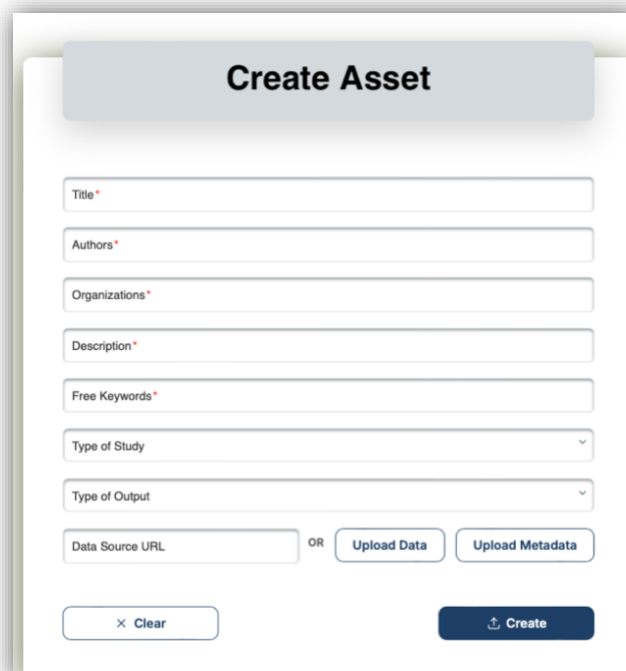



Figure 36: NIKH's Create asset form

The right tab of the Assets page (see Figure 37) named Federated Catalog shows the assets of the ecosystem according to the created policies and contracts. Each member is able to download an asset from the Data space ecosystem, if the policy of the contract is public, or request access if the policy of the contract is restricted. All the procedures regarding the arrangement of a contract and policies are described in the next subsection.


Assets

Federated Catalog

Title	Created By	Description	Date Created		Organization	Permission	Access
NextGEM: Next-Generation Integrated Sensing and Analytical System for Monitoring and Assessing Radiofrequency Electromagnetic Field Exposure and Health	npetro	The evolution of emerging technologies that use Radio Frequency Electromagnetic Field (RF-EMF) has increased the interest of the scientific community and society regarding the possible adverse effects...	2025-02-26	More »	FORTH	Open	Download
Towards the development of a network provisioning platform for data exchange in the health data space	npetro	The deployment of network provisioning mechanisms across data locations in the health domain requires a secure environment to enable reliable, trustworthy, and sovereign data exchange, addressing all ...	2025-02-26	More »	FORTH	Restricted	Request
Controllable Local Propagation Environment to Maximize the Multiplexing Capability of Massive MIMO Systems	schettino	The capability of controlling and modifying wireless propagation channels is one of the prerogatives of beyond-5G systems. In this paper, we propose the use of a controllable local propagation environ...	2025-02-26	More »	UCAS	Open	Download
A Simple and Low-Cost Technique for 5G Conservative Human Exposure Assessment	schettino	The purpose of this paper is to introduce a simple, low-cost methodology for estimating a conservative value of the maximum field level that can be radiated by a 5G base station useful for human expos...	2025-02-26	More »	UCAS	Restricted	Request



 Showing 1 - 4 of 4 results



 Page 1 of 1

← Previous




1

Next →

Figure 37: NIKH's Assets catalog view

7.2.2 Contracts

By selecting the Contracts tab on the Dataspace Ecosystem page (see Figure 33), a full member can visit the NIKH's Contracts page (see Figure 38). This is the page where all the contracts as well as the policies for the different owned asset is created.

Contracts and Policies			
Contracts		Policies	
Contract Name	Asset Name	Policy Name	
ljerph-public	ljerph20126085	no-restriction-policy	
Electronics-public	electronics12092022	no-restriction-policy	
CSCN-restricted	CSCN63674.2024.10849743	restricted-policy	

Showing 1 - 3 of 3 results

Page 1 of 1

← Previous 1 Next →

Create Contract

Figure 38: NIKH's Contracts view

The Contracts tab (left), provides information on the created contracts. Additionally, a trashcan symbol at the end of each presented contract entry allows authorized individuals (Supervisors or Full Members), to delete the particular contract from the premises. A “Create Contract” option is also available in the form of a button.

By clicking the “Create Contract” button (offered only to Supervisors and Full Members), the user invokes a form (see Figure 39) that contains all the relevant fields necessary for adding a new contract for selected assets.

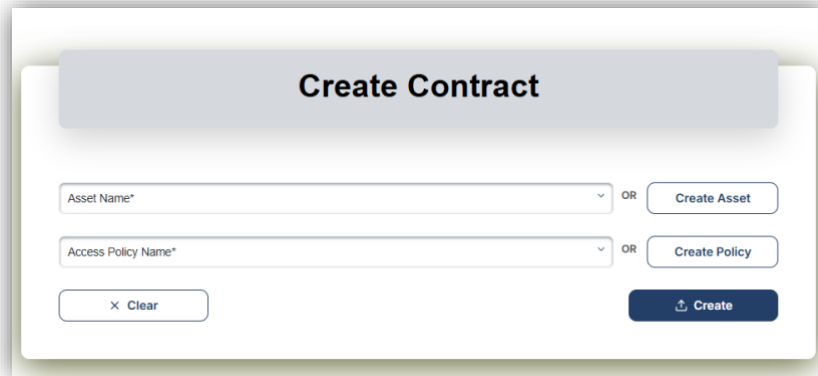


Figure 39: NIKH's Create contract form

The required fields are:

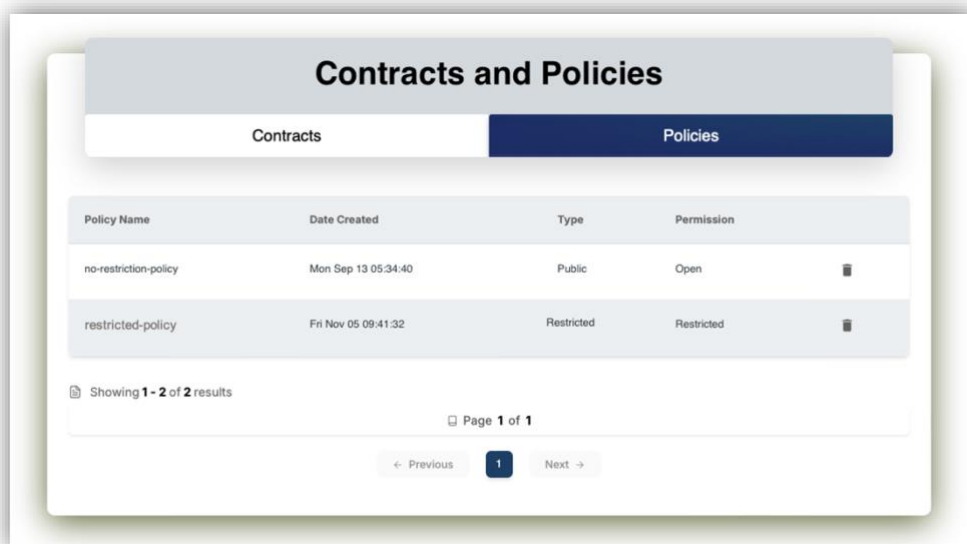
- Asset Name, selected from a drop-down list that shows all the assets belonging to the user’s premise
- Access Policy Name, selected from a drop-down list that shows all the policies belonging to the user’s premise



Alternatively, the “Create Contract” form offers a “Create Asset” button that opens the “Create Asset” form as described previously, and a “Create Policy” button that opens the “Create Policy” form as described in the next section.

7.2.3 Policies

The right tab (see Figure 40) on the Contracts page (see Figure 38), presents the policies that the member’s organization has created and are stored within the premises. Additionally, a trashcan symbol at the end of each presented policy entry allows authorized individuals (Supervisors or Full Members), to delete the particular policy from the premises. A “Create Policy” option is also available in the form of a button.

By clicking the “Create Policy” button (offered only to Supervisors and Full Members), the user invokes a form that contains all the relevant fields necessary for adding a new policy into the premises.



Policy Name	Date Created	Type	Permission	
no-restriction-policy	Mon Sep 13 05:34:40	Public	Open	
restricted-policy	Fri Nov 05 09:41:32	Restricted	Restricted	

Showing 1-2 of 2 results

Page 1 of 1

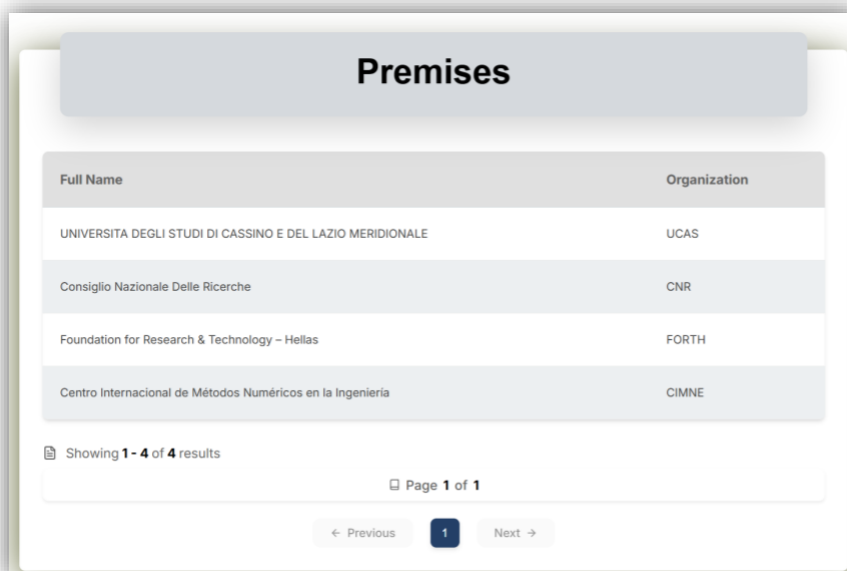
← Previous 1 Next →

Figure 40: NIKH's Policies view

7.2.4 Premises

By selecting the Premises tab on the Dataspace Ecosystem page (see Figure 33), the visiting member can see the registered premises in the NextGEM Ecosystem (see Figure 41). Each of the Premises within this view, indicates that presence of a Data space Connector, indicating that the listed organizations are active participants within NIKH. NIKH Premises/Connectors are created by NIKH's administrator/developers.

The full list of premises is offered only to users designated as Supervisors or Full Members. The rest of the users (with a Team Member role), will be able to see the premises they belong to.



Premises	
Full Name	Organization
UNIVERSITA DEGLI STUDI DI CASSINO E DEL LAZIO MERIDIONALE	UCAS
Consiglio Nazionale Delle Ricerche	CNR
Foundation for Research & Technology – Hellas	FORTH
Centro Internacional de Métodos Numéricos en la Ingeniería	CIMNE

Showing 1 - 4 of 4 results

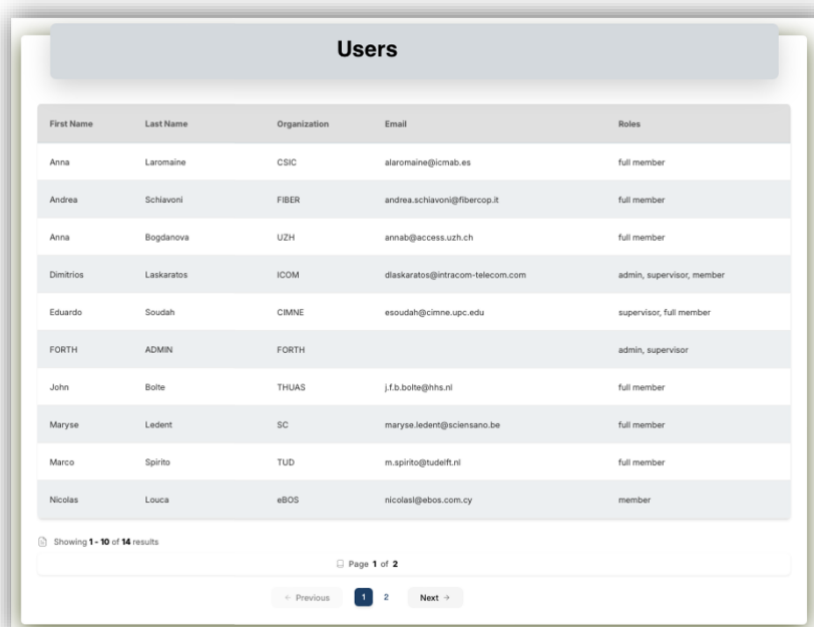
Page 1 of 1

← Previous 1 Next →

Figure 41: NIKH's Premises view

7.2.5 Users

Lastly, by selecting the Users tab on the Dataspace Ecosystem page (see Figure 33), the visiting member is presented with NIKH's Users page (see Figure 42). The Users page offers a view of the members available within the NIKH Dataspace Ecosystem. Additional useful information for each of the registered NIKH members is also presented. NIKH Users are created by NIKH's administrator/developers.



Users				
First Name	Last Name	Organization	Email	Roles
Anna	Laromaine	CSIC	alaromaine@icmab.es	full member
Andrea	Schiavoni	FIBER	andrea.schiavoni@fibercop.it	full member
Anna	Bogdanova	UZH	annab@access.uzh.ch	full member
Dimbrios	Laskaratos	ICOM	dlaskaratos@intracom-telecom.com	admin, supervisor, member
Eduardo	Soudah	CIMNE	esoudah@cimne.upc.edu	supervisor, full member
FORTH	ADMIN	FORTH		admin, supervisor
John	Boite	THUAS	j.f.b.boite@hhs.nl	full member
Maryse	Ledent	SC	maryse.ledent@sciensano.be	full member
Marco	Spirito	TUD	m.spirito@tudelft.nl	full member
Nicolas	Louca	eBOS	nicolas@ebos.com.cy	member

Showing 1 - 10 of 14 results

Page 1 of 2

← Previous 1 2 Next →

Figure 42: NIKH's Users view

8 Evaluation of NIKH Data space platform in realistic scenarios

To evaluate the practical application and adaptability of the proposed data space solution, we investigate its effectiveness, storage capabilities, and data governance based on a data-oriented approach on various deployment scenarios. These scenarios emphasize secure, flexible data exchange between stakeholders, addressing the unique data management needs of different organizational environments within the NextGEM consortium. Each scenario highlights the operational requirements, challenges, and essential functionalities necessary for effective deployment within a health data space ecosystem.

8.1 NextGEM Realistic Scenarios

To validate the applicability of the NIKH data space capabilities within the developed NIKH platform, a variety of realistic NextGEM usage scenarios are proposed, in which the functionalities of data space services are evaluated (Figure 43).

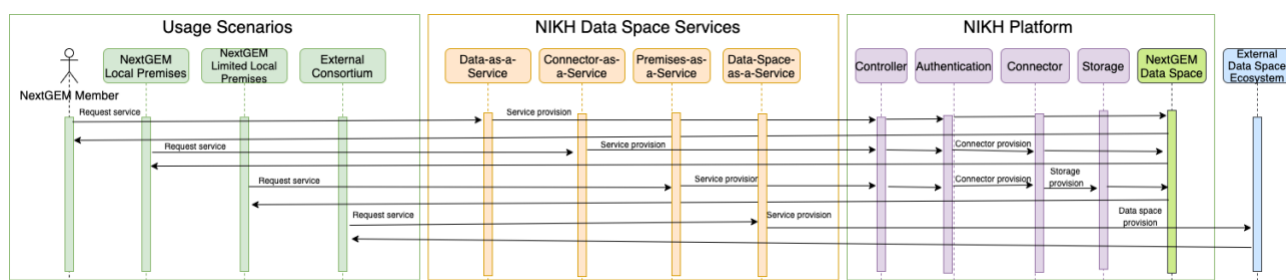


Figure 43: NIKH Platform data space services in NextGEM realistic usage scenarios

8.1.1 Usage Scenario 1 – NextGEM Members

The first usage scenario focuses on users within the NextGEM consortium who aim to retrieve data through a secure, permission-based platform. A NextGEM Member is any registered and authorized user on the NIKH platform, with access capabilities varying according to predefined permissions. To support these requirements, the Data-as-a-Service (DaaS) model enables seamless data access for users, eliminating the need for on-premises infrastructure or dedicated connectors for each NextGEM Member. Members are registered via the Keycloak access control mechanism to gain access to the available data through the NIKH controller. Access is granted through the centralized data space connector of the NIKH platform, which aggregates requested data from all network participants in compliance with data space protocols. This setup provides streamlined access and supports all operations without requiring additional data storage while ensuring high availability, scalability, and reliable access for NextGEM Members.

8.1.2 Usage Scenario 2 – NextGEM Local premises

The second usage scenario involves deploying necessary services on both the NIKH platform and the organization's infrastructure to facilitate data exchange between data providers and consumers. The local premises infrastructure ensures the protection of sensitive data, supporting seamless data transfer and sharing among all involved partners. This approach allows stakeholders to maintain secure and protected data within their own infrastructure while fostering efficient collaboration and sharing practices. In alignment with the proposed architecture, the Connector-as-a-Service (CaaS) model offers the deployment of a dedicated container running the EDC connector, along with additional containers running authority services, pre-deployed to support all local premises needs. A storage service is also deployed within the local infrastructure to enable persistent storage of scientific data. Consequently, this model demands sufficient storage capabilities on the local premises infrastructure, while data availability, scalability, and reliability depend on both the local premises infrastructure and the NIKH platform. Data isolation further ensures data governance, as organizations retain full control over their data.

8.1.3 Usage Scenario 3 – NextGEM Limited Local Premises

The third usage scenario involves entities deployed on the NIKH platform, enabling the exchange of scientific data from a data producer to other organizations through a shared infrastructure. Limited local premises utilize a common pool of resources that allow for the deployment of a dedicated connector and other necessary services for data collection and transfer, including an object storage service for storage needs. This approach ensures that

organizations using this model can maintain secure and protected data within the NIKH platform. In alignment with the proposed architecture, the Premises-as-a-Service (PaaS) model deploys a dedicated container running the EDC Connector, with essential services—such as authority services and a common object storage service for persistent storage—already deployed on the NIKH platform. Consequently, this model does not require individual storage capabilities, as the NIKH platform ensures high availability, scalability, and reliable data access. However, the centralized storage of data affects data governance, as organizations do not retain full control over their data space.

8.1.4 Usage Scenario 4 – External Data space ecosystem

The final usage scenario involves a group of entities or organizations, such as members of a project consortium or initiative, that need to deploy their own data space ecosystem to facilitate data exchange in alignment with data space principles. The Data-Space-as-a-Service (DSaaS) model allows for the creation of an external data space ecosystem tailored to the interests of these entities. Consequently, the service provider can replicate the proposed platform (including the Kubernetes cluster, controller, authority services, premises, storage, connectors, etc.) according to the specific needs and capabilities of the stakeholders involved. This approach enables the NIKH data space platform to be re-deployed in a centralized or remote location based on entities' requirements (e.g., number of participants, storage capabilities) to support the establishment of a fully functional data space ecosystem.

8.2 Discussion

As previously discussed, the NIKH platform provides data-sharing functionalities across various realistic scenarios, leveraging developed data space mechanisms to accommodate organizations with differing levels of infrastructure and needs. The primary goal of this Data Space topology is to enable controlled, sovereign, and secure data exchange and sharing among stakeholders. Within this framework, data sovereignty and trust are upheld, as each participant can apply usage restrictions to their data and monitor transactions through continuous tracking. Security is further reinforced through identity verification for each participant. Additionally, the framework supports metadata storage and query functionalities, allowing participants to search for relevant data sources and request access to specific datasets. In all scenarios, ecosystem members can access, publish, or acquire datasets under specified rules. Organizations or users wishing to utilize a dataset published within the ecosystem can request access, relying on data governance policies to validate the request under various conditions and locations. The different functionalities offered by NIKH data space services, along with the operational and technical requirements for end-users within these scenarios, are summarized in Table 8. Consequently, with standardized implementation procedures, the NIKH data space platform provides relevant stakeholders with an efficient and streamlined way to access the data space, update information, and provide data to involved parties.

Table 8: Validation of Data space services in realistic scenarios

Usage Scenario	Service	Storage	Availability	Scalability	Reliability	Data Governance	Isolation
US1- Members	Data-as-a-service	-	High (NIKH)	High (NIKH)	High (NIKH)	No	-
US2 - Local Premises	Connector-as-a-service	Depends on premises	Depends on premises	No	Depends on premises	Yes	Physical
US3 – Limited Local Premises	Premises-as-a-service	Limited (NIKH)	High (NIKH)	High (NIKH)	High (NIKH)	No	Virtual
US4 - External Data Space	Dataspace-as-a-service	Limited (NIKH)/ Depends on external space	High (NIKH)/ Depends on external space	High (NIKH)/ Depends on external space	High (NIKH)/ Depends on external space	Yes	Virtual / Physical

9 Conclusion

Deliverable D6.7 reports on the activities of Task 6.2 “Network provision and links with EU health data space”, which is part of WP6 on the “Development of NextGEM Innovation and Knowledge Hub”. To that end, the deliverable showcases the different implementation activities and frameworks utilized in the designing phase of NIKH’s components, establishing and ensuring a secure environment with respect to the legal/regulatory issues for health data sharing between the NextGEM. Within this scope, an overview of the platform’s services was provided, highlighting tailored data management strategies. Additionally, an investigation into the most efficient minimum viable data space mechanisms was conducted, identifying the most suitable one for the platform. The report outlines the network provision mechanisms between the distributed data locations and reports on the development status towards the development of the NIKH platform. Development details of the NIKH data space platform were also provided, along with an evaluation of its applicability under realistic scenarios. The next steps will include refining and expanding NIKH data space functionalities related to network provisioning among NextGEM partners, facilitating the exchange of experimental results across the various NextGEM case studies.

10 References

- [1] Menachemi, N., Rahurkar, S., Harle, C. A., & Vest, J. R. (2018). The benefits of health information exchange: an updated systematic review. *Journal of the American Medical Informatics Association: JAMIA*, 25(9), 1259–1265. <https://doi.org/10.1093/jamia/ocy035>
- [2] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. doi:10.1038/sdata.2016.18
- [3] Sawadogo, P., & Darmont, J. (2020). On data lake architectures and metadata management. *Journal of Intelligent Information Systems*, 56(1), 97–120. doi:10.1007/s10844-020-00608-7
- [4] Ramakrishnan, R., Sridharan, B., Douceur, J. R., Kasturi, P., Krishnamachari-Sampath, B., Krishnamoorthy, K., ... Venkatesan, R. (2017). Azure Data Lake Store: A Hyperscale Distributed File Service for Big Data Analytics. *Proceedings of the 2017 ACM International Conference on Management of Data*, 51–63. Presented at the Chicago, Illinois, USA. doi:10.1145/3035918.3056100
- [5] Sarramia, D., Claude, A., Ogereau, F., Mezhoud, J., & Mailhot, G. (2022). CEBA: A Data Lake for Data Sharing and Environmental Monitoring. *Sensors (Basel, Switzerland)*, 22(7), 2733. <https://doi.org/10.3390/s22072733>
- [6] Deligiannis, K., Raftopoulou, P., Tryfonopoulos, C., Platis, N., & Vassilakis, C. (2020). Hydria: An Online Data Lake for Multi-Faceted Analytics in the Cultural Heritage Domain. *Big Data and Cognitive Computing*, 4(2). doi:10.3390/bdcc4020007
- [7] Manco, C., Dolci, T., Azzalini, F., Barbierato, E., Gribaudo, M., & Tanca, L. (2023). HEALER: A Data Lake Architecture for Healthcare.
- [8] Wang, S., Zhang, Y., & Zhang, Y. (2018). A blockchain-based framework for data sharing with fine-grained access control in decentralized storage systems. *Ieee Access*, 6, 38437–38450.
- [9] Xia, Q. I., Sifah, E. B., Asamoah, K. O., Gao, J., Du, X., & Guizani, M. (2017). MeDShare: Trust-less medical data sharing among cloud service providers via blockchain. *IEEE access*, 5, 14757–14767.
- [10] Theodouli, A., Arakliotis, S., Moschou, K., Votis, K., & Tzovaras, D. (2018, August). On the design of a blockchain-based system to facilitate healthcare data sharing. In 2018 17th IEEE International Conference on Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE) (pp. 1374–1379). IEEE.
- [11] Shen, B., Guo, J., & Yang, Y. (2019). MedChain: Efficient healthcare data sharing via blockchain. *Applied sciences*, 9(6), 1207.
- [12] Xie, J., Yu, F. R., Huang, T., Xie, R., Liu, J., & Liu, Y. (2019). A Survey on the Scalability of Blockchain Systems. *IEEE Network*, 33(5), 166–173. doi:10.1109/MNET.001.1800290
- [13] Otto, B., Lohmann, S., Steinbuss, S., Teuscher, A., Auer, S., Boehmer, M., ... & Woerner, H. (2018). Ids reference architecture model. industrial data space. version 2.0.