DATASPACE 4HEALTH LUXEMBOURG

A Technical Reference Architecture to Interconnect the Care and Research Dataspaces

Document ID : DS4H_WP5_D5.3

Hee Kim, Maximilian Fünfgeld

Luxembourg Institute of Health (LIH)

Date: 30/07/2025

Funding

This project has received funding from the Ministry of the Economy of Grand Duchy of Luxembourg under Grant agreement no 20230505RDI170010392869.

Disclaimer

The documents published by the consortium are intended solely for informational purposes and reflect the views and opinions of the consortium members at the time of publication and within the scope of the Dataspace4Health project (DS4H).

The ideas expressed herein do not necessarily reflect the official policy or position of the funding entity.

Efforts have been made to achieve the relevance of the content; however, the consortium does not make any representation regarding its completeness and accuracy.

This document may be subject to further revision.

This document is intended to be published on the Dataspace4Health website.

Table of Versions

Version n°	Issue Date	Reason for change
0.0	02/07/2025	First draft.
0.1	09/07/2025	The internally reviewed version is shared with all partners.
0.2	16/07/2025	Pseudonymisation should be used instead of anonymisation. It is required for the oncology use case.
0.3	18/07/2025	The scope of this document is clarified with focus on interconnections with EHDS actors in chapter 1. The latest DS4H Reference Architecture is referenced in chapter 2. API specifications are indicated in chapter 3. Section 4.2 covers the archiving ETL process, instead of the Right-to-Erasure process.
1.0	30/07/2025	The revised version is ready. The final version has been sent to the coordinator.

TABLE OF CONTENTS

Table of figures 4
1. Introduction 5
2. Related Works 5
3. Technical Components 9
3.1. National dataset catalogues 9
3.2. Data access application 10
3.3. Secure processing environment 11
3.4. Transversal components 11
3.4.1. Authentication and authorisation infrastructure 12
3.4.2. Communication infrastructure 12
4. Infrastructure Specifications 13
4.1. Provisioning a secure processing environment 13
4.2. Storage architecture for data volume and access patterns 14
5. Conclusion 16
References 16

TABLE OF FIGURES

Figure 1 : Simplified HealthData@EU architecture to interconnect care and research data spaces proposed by TEHDAS [3]. 6
Figure 2 : Reference Architecture proposed by the Dataspace4Health project WP3. 7

TABLE OF TABLES

Table 1. Actors in the HealthData@EU architecture and mapping to the DS4H architecture. 8
Table 2. Security requirements for communication between actors when the care and research data spaces are interconnected. 11
Table 3. Recommendation for resource planning for SPEs. 13
Table 4. Three storage architecture types: data lakes, data warehouses, and data marts. 14

Glossary

Abbreviation	Expression
AAI	Authentication and Authorization Infrastructure
AI	Artificial Intelligence

CISPE	Cloud Infrastructure Services Providers in Europe
DCAT	Data Catalogue Vocabulary
DCAT-AP	DCAT Application Profile
EHDS	European Health Data Space
ETL	Extract, Transform, Load
EUCS	EU Cybersecurity Certification Scheme
FHIR	Fast Healthcare Interoperability Resources
GDPR	General Data Protection Regulation
HDAB	Health Data Access Body
HealthDCAT-AP	HealthDCAT Application Profile
HPC	High-Performance Computing IdP Identity Provider
IPMS	Identifier-Matching and Pseudonym Management Service
ISO	International Standards Organisation
ISMS	Information Security Management System
LNDS	Luxembourg National Data Service
MTB	Molecular Tumor Board
NCP	National Contact Point
OAuth	Open Authorization
OIDC	OpenID Connect
OMOP	Observational Medical Outcomes Partnership
PDP	Policy Decision Point
PEP	Policy Enforcement Point
RBAC	Role-Based Access Control
SPE	Secure Processing Environment
SP	Service Provider
SSL	Secure Sockets Layer
SSO	Single Sign-On
TCP	Transmission Control Protocol
TEHDAS	Towards the European Health Data Space
TLS	Transport Layer Security
WGS	Whole-Genome Sequencing
W3C	World Wide Web Consortium

1. INTRODUCTION

The European Health Data Space (EHDS) [1] is a regulation at the European level to create a unified, secure, and interoperable data infrastructure for health data sharing across Europe. The EHDS addresses numerous challenges in the current health data domain. Data fragmentation persists across institutional and national boundaries, creating barriers that limit the potential for cross-institutional and cross-border healthcare services and collaborative research. Regulatory complexity emerges from the implementation of the GDPR [2] across member states. Technical heterogeneity, such as a lack of a common health data model and inconsistent infrastructure, hinders effective data sharing.

This document aims to address the first challenge - data silo by providing a blueprint for interconnecting care and research data spaces focused on the secondary use of health data for research purposes and innovation. It has been built upon the work presented in Deliverable 7.2 Options for the services and services architecture and infrastructure for secondary use of data in the EHDS [3] from the TEHDAS project. Developers can use this document as a guideline to understand the high-level design patterns of their relationships, with a particular focus on the key components outlined in Section 3.

There are two associated Deliverables within the Dataspace4Health (DS4H) project [4]. This document serves as a technical overview to facilitate communication among participants or entities in the EHDS. In contrast, Deliverable 3.1 [5] proposed the DS4H Reference Architecture, providing a comprehensive blueprint, while the technical specifications, such as API signatures, will be defined in the upcoming Deliverable 8.1 [6]. Both Deliverables provide technical references for implementing health dataspaces in line with Gaia-X principles. Gaia-X is a broader European initiative on federated, sovereign data infrastructures, launched in 2019, which defines a robust trust framework, open standards, and verified digital clearing services for secure, transparent, and interoperable data exchange [7].

2. RELATED WORKS

TEHDAS proposed a reference architecture to facilitate the secondary use of health data. This architecture defines key actors and their roles, and is designed to ensure secure access to health data. Figure 1 presents a simplified HealthData@EU architecture taken from the TEHDAS Deliverable 7.2, showing key actors and their interactions. The scope of the Dataspace4Health project is to connect domestic participants in Luxembourg, as indicated by the blue rectangular box in Figure 1. Nonetheless, this document addresses interactions with all entities, including Health data access body (HDAB), National Contact Point (NCP), and the EU Core Platform [8]. These interactions are essential to facilitate the secondary use of health data.

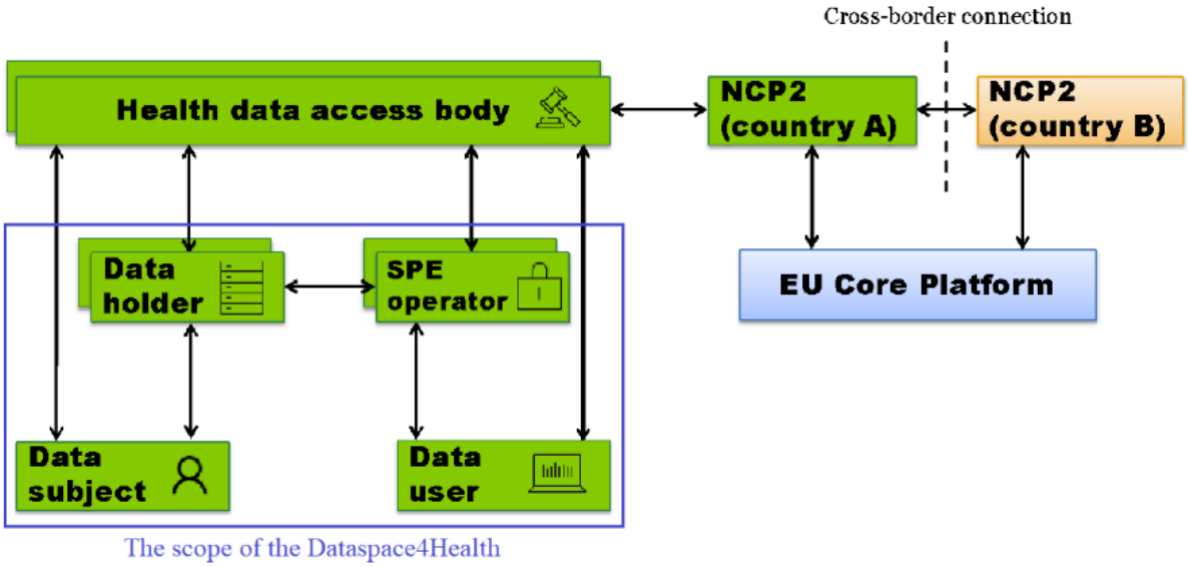

Figure 1 : Simplified HealthData@EU architecture to interconnect care and research data spaces proposed by TEHDAS [3].

The list below is the key actors of the HealthData@EU architecture and their responsibilities. The first five are the actors required to interconnect care and research data spaces domestically, while the rest are relevant to interconnect care and research data spaces cross-border.

- **Data subject** is the individual whose health data is being processed. Their rights, consent, and privacy are central to the entire architecture.
- **Data holder** is the entity that originally collects and stores the health data (e.g., hospitals, clinics, research institutions).
- **Data user** is the individual or legal entity requesting access to health data for secondary purposes (e.g., researchers, pharmaceutical companies).
- **Secure Processing Environment (SPE) operator** is the entity responsible for developing, maintaining, and operating the secure virtual environments where sensitive health data can be processed by the data user.
- **Health Data Access Body (HDAB)** is the national entity responsible for granting access to health data for secondary use. They act as trusted intermediaries, ensuring compliance with national regulations and data governance frameworks.
- **National Contact Point (NCP)** facilitates cross-border data access requests and acts as a liaison between HDABs in different member states and the EU Core Platform. NCP for secondary use can be the same as the coordinator Health Data Access Bodies (HDABs), stated in Article 75, paragraph 1 of the EHDS Regulation.
- **EU Core Platform** [8] is the central component facilitating the coordination and interoperability across national data spaces. It manages the EU Dataset Catalogue, routes data access applications, and supports cross-border data exchange. The management of data access requests might be done through this central platform.

On the other hand, the Dataspace4Health (DS4H) project [4] proposed a reference architecture [5] shown in Figure 2. It consists of five layers, each serving different functions. It is a technical architecture ensuring a federated, sovereign, and interoperable data space to interconnect healthcare providers and research institutes in Luxembourg.

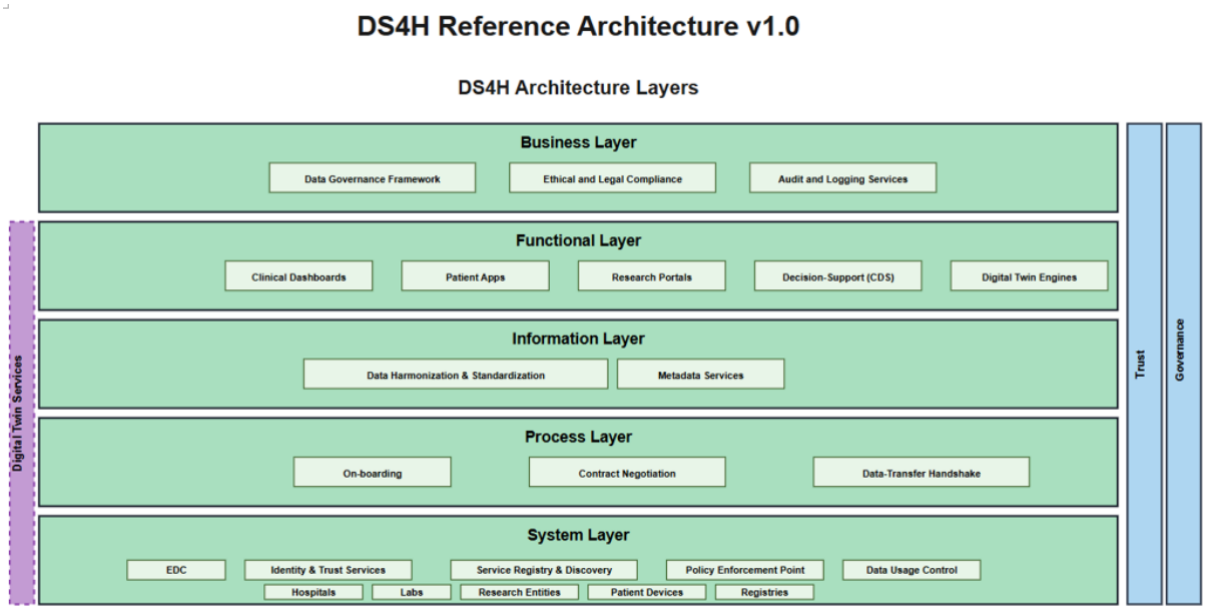

Figure 2 : Reference Architecture proposed by the Dataspace4Health project WP3.

The breakdown of each layer and the Gaia-X [7] relevance are presented as follows:

- **System Layer** enables secure, sovereign, and policy-compliant data exchange across participants in the data space. It provides the Gaia-X-compliant building blocks (e.g., EDC, identity, catalogue) that enable sovereign, policy-driven data sharing across federated entities. It integrates data sources including electronic health record systems, medical registries, biobanks, and research databases.
- **Process Layer** handles intelligent automation and coordination, running things like digital twin engines and tools that support clinical decisions. This layer makes sure workflows are dynamic, respect consent, and can seamlessly connect multiple institutions.
- **Information Layer** is responsible for enabling machine-readable and context-aware exchange of data among participants in the data space. This layer enables interoperability at scale while reducing ambiguity and misinterpretation across institutions and systems, as well as supports federated queries and AI readiness.
- **Functional Layer** refers to how users interact with the ecosystem. It considers interfaces and applications that support interaction by clinicians, patients, researchers, and administrators. Clinical dashboards, patient apps, and research portals are examples.
- **Business Layer** provides the legal, ethical, and operational framework that ensures data is used safely, fairly, and lawfully. It defines who does what, under which rules, and ensures trust, ethics, and compliance. Informed consent, secondary use authorization, opt-out mechanisms, and immutable logs should be implemented in this layer.

Table 1 presents the mapping between the HealthData@EU architecture and the DS4H general architecture. This table demonstrates the alignment between the general architecture proposed by the DS4H project with the HealthData@EU framework by linking actors from the HealthData@EU architecture to the corresponding layers of the DS4H architecture.

Table 1. Actors in the HealthData@EU architecture and mapping to the DS4H architecture.

HealthData@EU Actor	Mapping to the DS4H architecture
Data subject	Interacts via the **Functional Layer** (e.g., through consent management portals) and is the source of data in the **System Layer**.
Data holder	Resides primarily in the **System Layer** and **Process Layer** through data access portals.
Data user	Interacts primarily with the **Functional Layer** and **Process Layer** through data access portals.
SPE operator	Operates within the **System Layer** for providing a secure computational environment and the **AI & Functional Layer** for hosting analytical tools.
Health Data Access Body	Operates within the **Business Layer** for policy enforcement, access control decisions, and the **System Layer** for managing data access requests and potentially operating secure environments.
National Contact Point	Operates within the **Business Layer** and the **System Layer** for routing and coordinating requests.
EU Core Platform	It spans across the **Functional Layer** for web portal, **System Layer** for routing, catalogue management, **Information Layer** for metadata harmonization, and **Business Layer** for overarching policy enforcement.

3. TECHNICAL COMPONENTS

The main actors of the EHDS for interconnecting care and research data spaces are described in Section 2; however, it does not include the technical components needed in setting up and maintaining the infrastructure. This section focuses on technical specifications for three key technical components of the HealthData@EU infrastructure, namely: (1) the information systems to manage the national metadata catalogues; (2) the information systems to manage the cross-border data access applications and data requests; and (3) the secure processing environments (SPE). The guidelines for three components of the HealthData@EU are the outcome of a large effort of the TEHDAS WP7 activities on integrating the request made by the European Commission. The context and full guidelines are available in Annexes A to C [9]. These three core components are underpinned by transversal technical components like Authentication and Authorisation (e.g., eIDAS, OAuth 2.0 bearer tokens), and secure communication protocols (e.g., HTTPS/TLS).

The HealthData@EU Central Platform is scheduled to release a new version every four months, each accompanied by technical documentation for developers. The document includes architecture models and API specifications for the technical components. In the current Version 4 documentation [19], Section 8 details the API specifications. The final version of the platform is set to release in September 2027, with full operational status targeted for 2028, in line with the EHDS Regulation.

3.1. NATIONAL DATASET CATALOGUES

National dataset catalogues are one of the foundational pillars in the EHDS for enabling the discoverability of health data. It is anticipated that multiple data holders will be connected to a single Health Data Access Body (HDAB) per country. The HDAB manages the national dataset catalogue in a federated manner and promotes the adoption of the same standard among Data holders. And then,

the EU Core Platform will harvest these national dataset catalogues from each state member's HDAB to generate a comprehensive Catalogue at the EU level. This federated approach ensures that data remains under national governance while being discoverable at a European level.

HealthDCAT-AP[1] is the standardized health metadata in the EHDS to accommodate health-specific requirements proposed by TEHDAS Deliverable 6.2 [ref]. It is an extension of Data Catalog Vocabulary (DCAT) Application Profile (DCAT-AP)[2] for data portals in Europe promoted by the European Commission, built upon the DCAT[3] developed by the W3C.

- Considerations for data connector developers:
  - **Metadata publication**: In order to publish metadata to a national catalogue, it must adhere to the HealthDCAT-AP standard for creating and updating metadata entries, understanding required fields, controlled vocabularies, and versioning.
  - **API for data discovery**: Implement clients for the catalogue's RESTful APIs. This involves sending HTTP requests (GET, POST) with appropriate query parameters to search for datasets based on criteria like keywords, data types, and more.
  - **HealthDCAT-AP parsing**: Interpret the HealthDCAT-AP metadata schema. This ensures that the connector can identify dataset attributes, data formats, access points, and associated legal/governance information.

3.2. DATA ACCESS APPLICATION

The data access application will operate as a hybrid system designed to manage services in both centralised and decentralised manners. In essence, the system will support centralised data request handling and distributed grant services. Specifically, the EU Core Platform will facilitate the submission of data access requests from data users across Europe, enabling data users to submit data access applications. Meanwhile, the actual granting of data access will be given by national HDABs, each of which will operate its dedicated instance of an application management system.

This hybrid system allows customisation in the approval process per HDAB; however, it also introduces higher complexity in maintaining consistency across the system. For example, the EU Core Platform will directly receive data access applications containing the necessary information, whether these requests originate from an EU-level or national-level portal. It will then distribute these requests (data access applications) to the relevant HDABs that hold connections to the requested datasets or data statistics of the relevant data holder. Subsequently, the EU Core Platform will receive the decisions (approvals or rejections) taken by the HDABs and efficiently distribute them back to the original requesters.

By centralizing the coordination of requests while preserving national autonomy over access decisions, the EU Core Platform streamlines the overall application process and ensures efficient, structured routing of requests.

- Considerations for data connector developers:
  - **API for application submission:** Data users shall interact with the EU Core Platform through connectors with APIs.
  - **Status monitoring:** Implement polling or webhook listeners to receive updates on the status of submitted data access applications (e.g., pending, approved, rejected).

[1] https://healthdcat-ap.github.io/
[2] https://interoperable-europe.ec.europa.eu/collection/semic-support-centre/dcat-ap
[3] https://www.w3.org/TR/vocab-dcat-3/

Message queue tools such as Kafka[4] or RabbitMQ[5] are the candidate tools for asynchronous notifications of application status changes.

3.3. SECURE PROCESSING ENVIRONMENT

The secure processing environment (SPE) is a critical component of the European Health Data Space (EHDS) architecture, ensuring the secure access, compliant, and controlled process of sensitive health data for secondary purposes. SPEs provide a controlled, isolated, and secure environment for data users and enable data users to perform healthcare analyses without directly retrieving the source datasets. This idea aligns with the "code-to-data" paradigm, also a cornerstone of federated data processing, which minimizes data movement and significantly reduces the risk of data breaches or unauthorized access.

The SPE should ensure compliance with strict information security and data protection standards, including frameworks such as International Standards Organisation (ISO) 27001[6], the EU Cybersecurity Certification Scheme (EUCS)[7], the Five Safes framework [10], and the Data Protection Code of Conduct [11] by CISPE. These frameworks or schemes provide an overview of information security management systems (ISMS) and the consistency of SPE development across jurisdictions.

- Considerations for data connector developers:
  - **Common data modelling:** Raw health data must be pre-processed and transformed into the common data models (e.g., FHIR[8], OMOP[9]). Connectors might involve developing auto-ETL pipelines that handle data cleaning, mapping, and validation before ingestion.
  - **Data pseudonymization:** Connectors need to integrate with or perform pseudonymization methods before data enters the SPE. In Luxembourg, this service is delivered as a fully independent Trusted Third-Party (TTP) function co-created by Luxembourg National Data Service (LNDS)[10] and INCERT[11]. Identifier-Matching and Pseudonym Management Service (IPMS) for protecting sensitive personal data is delivered on an on-demand and project-specific basis.
  - **Output retrieval APIs:** Connectors should provide secure mechanisms for data users to retrieve approved outputs (e.g., aggregated results, statistical reports) from the SPE, ensuring that only non-identifiable data leaves the SPE. The approval mechanism that interacts with HDAB might also be implemented.
  - **API for SPE management:** Connectors might interact with SPE management APIs for provisioning compute resources, deploying analytical environments, or monitoring job status and security policies.

3.4. TRANSVERSAL COMPONENTS

[4] https://kafka.apache.org/
[5] https://www.rabbitmq.com/
[6] https://www.iso.org/standard/73906.html
[7] https://www.enisa.europa.eu/publications/eucs-cloud-service-scheme
[8] https://hl7.org/fhir/
[9] https://www.ohdsi.org/data-standardization/
[10] https://www.lnds.lu/
[11] https://www.incert.lu/

3.4.1. AUTHENTICATION AND AUTHORISATION INFRASTRUCTURE

Authentication and authorization infrastructure (AAI) is one of the foundational services for easing the user experience in the EHDS ecosystem, where participants need to be involved in interactions between various stakeholders and technological systems. AII includes user identity verification, credential management, and granular access control. Minimizing the complexity of user management across all systems is crucial for both seamless integration of the user journey and robust security.

A robust AAI will provide a consistent and uniform experience across all services, eliminating the need for users to manage multiple credentials for different components for every phase of the user journey. According to the TEHDAS Deliverable 7.2, it would be preferable for these AAI to be coordinated by member states first, and then coordinated by the EU Core Platform, while third-party operated AAI is not considered due to potential security risks. The ultimate goal is expected to enable a cross-border Single Sign-On (SSO) [12].

- Considerations for data connector developers:
  - **OAuth 2.0/OpenID Connect (OIDC):** Connectors should implement clients for OAuth 2.0 [13] and OpenID Connect [14]. This includes obtaining ID tokens, access tokens, and refresh tokens, handling token expiration, and storing tokens securely.
  - **eIDAS:**
    - The eIDAS framework [15] is a key enabler for secure cross-border transactions in Europe, and the Gaia-X Trust Framework incorporates its principles for electronic identification, authentication, and trust services.
    - Connectors need to be registered as a Service Provider (SP) in the eIDAS network. This process is handled through each member state's eIDAS Single Point of Contact, involving strict security audits and legal compliance checks.
    - Integrate the registered connector with the eIDAS Node for communication between national Identity Providers (IdPs) and the SP (connector) using OIDC.
  - **Policy enforcement point (PEP):** Integrate with Policy Enforcement Points (PEPs) that interact with policy decision points (PDPs). This means making authorization requests to the AAI before accessing data or performing operations.
  - **Role-Based Access Control (RBAC):** Understand how to leverage attributes (e.g., user roles, data sensitivity, purpose of use) provided by the AAI to make fine-grained authorization decisions within the connector's logic.
  - Keycloak[12] and Keyrock-FIWARE[13] are powerful platforms, integrating eIDAS authentication on top of robust existing features such as user management, Single Sign-On (SSO), and comprehensive authorization capabilities, including Role-Based Access Control (RBAC).

3.4.2. COMMUNICATION INFRASTRUCTURE

High-bandwidth, low-latency, and secure network connectivity are expected for interconnecting the care and research data spaces. Communication requirements are different based on the participating actors and data type. Table 2 refined the content proposed by the TEHDAS Deliverable 7.2, which is chronologically structured from publishing metadata by the data holder to data analysis by the data user.

Table 2. Security requirements for communication between actors when the care and research data spaces are interconnected.

[12] https://www.keycloak.org/
[13] https://keyrock-fiware.github.io/

Step	Actor					Data volume		Security level	
	Data holder	Data user	SPE	HDAB	Central platform	Small	Big	High	Highest
Publish metadata for catalogue	X			X		X		X	
Data access request		X		X	X	X			X
Grant data permit		X		X	X	X			X
Transfer Pseudonymised or aggregated data	X		X				X	X	
Transfer pseudonymised data	X		X				X		X
Data analysis		X	X			X			X
Federated analysis			X			X			X
Incidental results			X	X		X			X

For high-bandwidth data transfers (e.g., high-resolution imaging data), the communication channels between the SPE and data holders might be non-TCP protocols, such as those used by Aspera[14], while the other communication links can use standard TCP connections. In all cases, data must be secure and maintain integrity utilizing either network-layer encryption (e.g., transport-level encryption such as SSL/TLS [16]).

4. INFRASTRUCTURE SPECIFICATIONS

For data connector developers, understanding how to provision the computational resource optimally is crucial to handle diverse data types and analytical workloads. This section elaborates on how to provision for the computational and storage infrastructure, differentiating recommendations based on data characteristics and application types from lightweight text analysis to intensive omics data processing.

4.1. PROVISIONING A SECURE PROCESSING ENVIRONMENT

The specific hardware provisioning within an SPE should be dynamically configured to meet the demands of different data types and analytical methods. For instance, High-Performance Computing (HPC) systems are anticipated for intensive tasks like omics-related analysis, while GPU-powered solutions are expected for AI-associated analysis, where parallel processing is crucial. Basic work stations are sufficient for regular statistical inference.

For generic workloads for small data size, providing at least 2 CPU cores (x86, Intel or AMD) and 12 GB RAM is recommended, as these specifications align with popular cloud-based platforms, Google Colab free tier. It might be a starting point to conduct a small to medium-sized data analysis. For the medium to large data, we reference the hardware allocated for Galaxy platform[15] users. A registered Galaxy user has a 250 GB disk space quota, while each user can execute a maximum of 6 concurrent

[14] https://download.asperasoft.com/download/docs/csrv/3.3.4/linux/html
[15] https://galaxyproject.org

jobs, each is assigned 16 processing cores [17]. The Galaxy project provides a cloud platform for tens of thousands of scientists across the world to analyze large biomedical datasets. Highly intensive workloads such as whole-genome sequencing (WGS) require high-performance computing (HPC) because it involves processing terabytes of raw data, running complex alignment algorithms. The specifications of one of the European HPC Supercomputers [18] provides more powerful hardware for large-scale, highly parallel, or GPU-required workload. All recommendations for three scenarios are summarized in Table 3. If SPE accommodates more data users simultaneously, this recommendation should be scaled up (e.g., by multiplying the per-user requirements).

Table 3. Recommendation for resource planning for SPEs.

Data size	Small to Medium	Medium to Large	Large to Very Large
Workload	LLM API calls, statistical analysis	Image data analysis, variant analysis, metagenomics, proteomics, and transcriptomics	Whole-genome sequencing (WGS) of many samples
Processor	2 CPUs +	Up to 16 cores per job	48+ cores per node
GPU	None or mid-range GPU (e.g., T4 or K80)	High-end GPUs (e.g., V100 / A100) limited availability	4 high-end GPUs (e.g., A100) per node
Storage	~100 GB	250 GB per user	multi-PB capacity

4.2. STORAGE ARCHITECTURE FOR DATA VOLUME AND ACCESS PATTERNS

The design of the storage infrastructure must account for the primary purpose of storing data, data formats, and the varied access patterns required by different analytical workflows. Modern approaches to optimize data management are classified into three storage architecture types as presented in Table 4.

Persistent systems for data storage require robust lineage tracking and purge mechanisms to ensure downstream erasure requests are honored. However, given that the SPE is designed as a transient workspace where all data is purged after processing, Right-to-Erasure (lineage metadata and purge hooks) becomes redundant. All data in the SPE will be deleted after task completion. However, ETL pipelines constructing data lakes to data marts are required. In all cases, the process of validation and archival of exemplar data for reproducibility is related to the SPE where the analyses were performed. According to Section 5.3.2 of TEHDAS Deliverable 7.2 [3], the preparation of the results for a possible cataloguing and archiving purpose requires external repositories, such as Zenodo, EU Open Data Portal, EOSC, or the European Health Information Portal.

Table 4. Three storage architecture types: data lakes, data warehouses, and data marts.

Storage type	Data lakes	Data warehouses	Data marts
Primary purpose	Massive, centralized repository for raw data; Archiving.	Structured repository for routinely generated data; Reporting, business intelligence.	Data subsets for specific projects; Individual data requests for longer access periods.
Data type	Multi-modal raw data (binary, structured, unstructured)	Structured, routinely updated	Curated subset of data (structured or semi-structured)
Access pattern	Store and access (schema-on-read).	ETL pipeline (schema-on-write); Fast query performance on structured data; Supports routine analyses	Provides tailored data for specific research needs; Supports replication of analyses

5. CONCLUSION

This Technical Reference Architecture serves as a roadmap for technical teams to understand the overarching vision and technical components of the EHDS that extend the development of data connectors. It provides a comprehensive and practical framework essential for the realization of interconnected care and research data spaces within the EHDS. The reference architecture pays particular attention to the integration of its three core EHDS components - national metadata catalogues, data access applications, and secure processing environments - with data connectors combining rigorous regulatory compliance with technical guidelines.

REFERENCES

1. European Union. (2025, February 11). Regulation (EU) 2025/327 of the European Parliament and of the Council on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847 (Text with EEA relevance). Official Journal of the European Union, L 327. Retrieved March 21, 2025, from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202500327

2. GDPR (General Data Protection Regulation): European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union, L119, 1–88. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679

3. TEHDAS project: Deliverable 7.2: options for the services and services architecture and infrastructure for secondary use of data in the EHDS.2023; Retrieved July 1, 2025, from https://tehdas.eu/app/uploads/2023/07/tehdas-options-for-the-services-and-services-architecture-and-infrastructure.pdf

4. Dataspace4Health, www.dataspace4health.lu. Accessed July 2025.

5. Deliverable 3.1 (from Dataspace4Health project)

6. Deliverable 8.1 (from Dataspace4Health project)

7. Gaia-X. Gaia-X Association, 2025, gaia-x.eu. Accessed 11 July 2025.

8. EU Core Platform, https://acceptance.data.health.europa.eu/healthdata-central-platform/home?locale=en, Accessed July 2025.

9. Joint Action Towards the European Health Data Space (TEHDAS). Options for the Services and Services' Architecture and Infrastructure for Secondary Use of Data in the EHDS. TEHDAS, July 2023.

10. Soiland-Reyes, Stian, et al. "Five Safes RO-Crate: FAIR Digital Objects for Trusted Research Environments." International FAIR Digital Objects Implementation Summit 2024. TIB Open Publishing, 2024.

11. Cloud Infrastructure Service Providers in Europe (CISPE). Data Protection Code of Conduct for Cloud Infrastructure Service Providers under GDPR. 9 Feb. 2021.

12. De Clercq, Jan. "Single sign-on architectures." International Conference on Infrastructure Security. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002.

13. Hardt, Dick. The OAuth 2.0 authorization framework. No. rfc6749. 2012.

14. Sakimura, Nat, et al. "OpenID Connect Core 1.0 incorporating errata set 1." The OpenID Foundation, specification 335 (2014).

15. Regulation (EU) No. 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC. Official Journal of the European Union L 257 (28 Aug. 2014): 73–114. 1 July 2016.

16. Rescorla, Eric. "The Transport Layer Security (TLS) Protocol Version 1.3." RFC 8446, Aug. 2018. Internet Engineering Task Force.

17. Langmead, Ben, and Abhinav Nellore. "Cloud computing for genomic data analysis and collaboration." Nature Reviews Genetics 19.4 (2018): 208-219.

18. Wellein et al., JUWELS: Modular Tier-0/1 Supercomputer at the Jülich Supercomputing Centre, 2020 ISC High Performance Conference

19. HealthData@EU central platform - Publications Office of the EU, https://op.europa.eu/en/publication-detail/-/publication/d2947362-3c36-11f0-8a44-01aa75ed71a1