

Viewpoint

# Reality Check: The Aspirations of the European Health Data Space Amidst Challenges in Decentralized Data Analysis

Holger Fröhlich<sup>1</sup>, PhD; Anne Funck Hansen<sup>2</sup>, PhD; Mika Hilvo<sup>3</sup>, PhD; Gunther Jansen<sup>4</sup>, PhD; Sumit Madan<sup>1</sup>; Sobhan Moazemi<sup>1</sup>, PhD; Sanziana Negreanu<sup>5</sup>; Venkata Satagopam<sup>6</sup>, PhD; Phil Gribbon<sup>7</sup>, PhD; Christian Muehldyck<sup>5</sup>, PhD

<sup>1</sup>Fraunhofer Institute for Algorithms and Scientific Computing, Sankt Augustin, Germany

<sup>2</sup>Fraunhofer Society for the Advancement of Applied Research, Munich, Germany

<sup>3</sup>VTT Technical Research Centre of Finland, Espoo, Finland

<sup>4</sup>Novartis (Switzerland), Basel, Switzerland

<sup>5</sup>Johnson & Johnson (Germany), Norderstedt, Germany

<sup>6</sup>Luxembourg Centre for Systems Biomedicine, Belvaux, Luxembourg

<sup>7</sup>Fraunhofer Institute for Translational Medicine and Pharmacology, Hamburg, Germany

**Corresponding Author:**

Holger Fröhlich, PhD  
Fraunhofer Institute for Algorithms and Scientific Computing  
Schloss Birlinghoven  
Sankt Augustin 53757  
Germany  
Phone: 49 151 7059794  
Email: [holger.froehlich@scai.fraunhofer.de](mailto:holger.froehlich@scai.fraunhofer.de)

## Abstract

The European Health Data Space (EHDS) aspires to enable secure, interoperable, and decentralized health data usage across Europe. This paper explores legal and technical challenges in implementing EHDS goals, particularly for secondary data use. It highlights federated and swarm learning as promising yet complex solutions, requiring robust infrastructure, standardization, and regulatory clarity. We emphasize the need for coordinated legislative and technological advances to realize EHDS ambitions.

*J Med Internet Res* 2025;27:e76491; doi: [10.2196/76491](https://doi.org/10.2196/76491)

**Keywords:** European Health Data Space; federated learning; swarm learning; General Data Protection Rule; federated data analysis

## Introduction

### Background

The European Health Data Space (EHDS) aims at transforming health care delivery, innovation, and research across Europe [1]. The EHDS is based on a legal framework on which the European Parliament and the Council reached a political agreement in spring 2024. It has the following three overarching goals: control over personal health data, developing a market for electronic health records (EHRs), and facilitating secondary data use.

### Control Over Personal Health Data

This goal is to ensure that individuals can access and manage their own health data securely, which is at the core of the

EHDS vision. By enabling portability of data across the European Union (EU), the initiative seeks to empower the citizens and health care providers to make better-informed decisions. Notably, member states can offer a complete drop-out to their citizens.

### Developing a Market for EHRs

The EHDS seeks to initiate a competitive market for EHR systems that are accessible, efficient, secure, and interoperable across member states.

### Facilitating Secondary Data Use

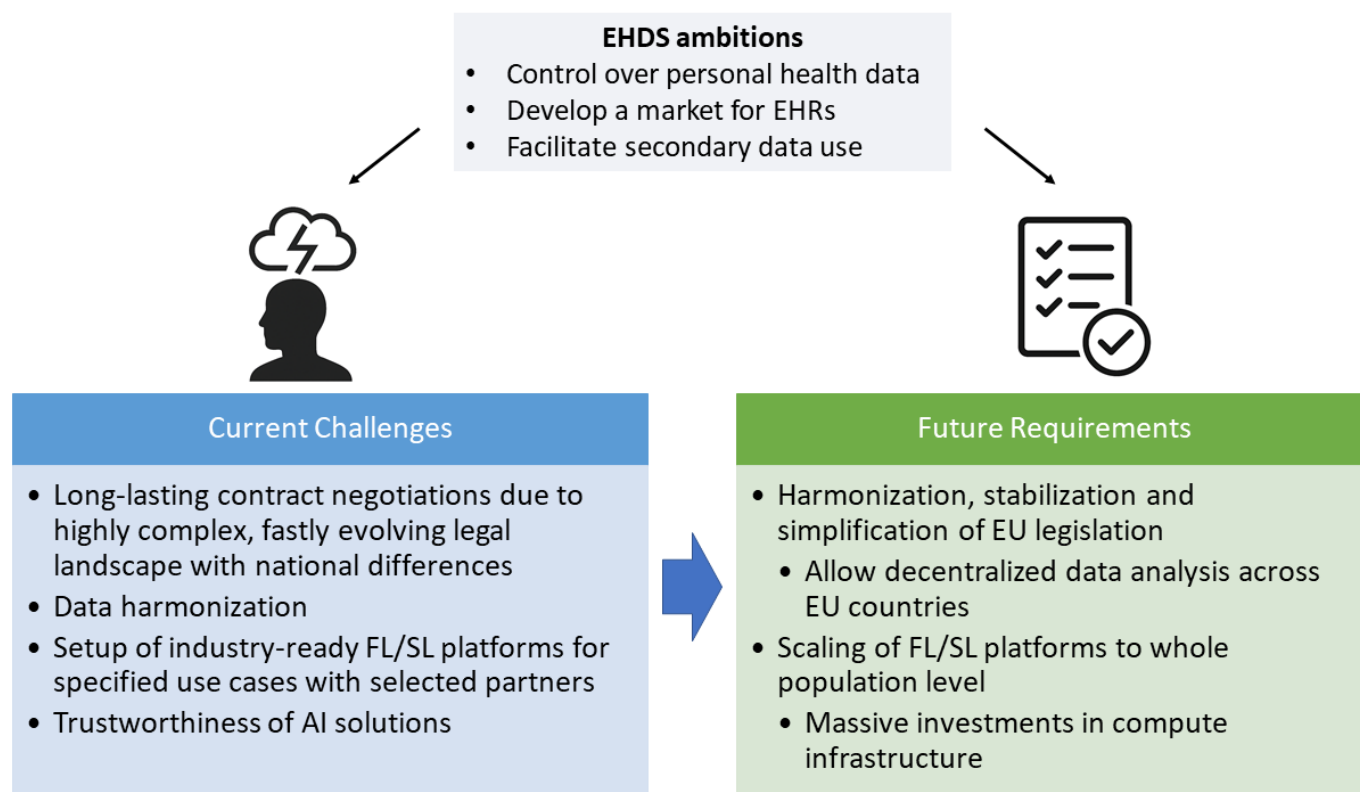
The EHDS aims to not only break down silos that hinder the flow of data between stakeholders in the health care system, but also promote the secondary use of these data for research, innovation, policy-making, and regulatory activities.

The EHDS primary legislation provides an overarching framework but does not provide any technical infrastructure or implementation details [2]. However, the secondary legislation and guidance documents from ongoing initiatives such as TEHDAS2 [3] are working on further details to support the implementation in the EU member states.

## Challenges

Projects aiming at implementing a technical infrastructure compatible with the secondary legislation ambitions of the EHDS are currently confronted with several legal as well as technical challenges (Figure 1).

**Figure 1.** Current challenges and future requirements for a successful realization of the European Health Data Space (EHDS). AI: artificial intelligence; EHR: electronic health record; EU: European Union; FL: federated machine learning; SL: swarm learning.



## Legal Challenges

The ambition of the EHDS to establish a market for EHRs and to facilitate the secondary use of health data is challenging considering the current heterogeneous interpretation of the General Data Protection Regulation (GDPR) regarding the use of sensitive patient-level data [4]. While processing of health data without the explicit consent of patients is principally possible for research purposes within the GDPR, this typically requires careful documentation, risk assessments, and compliance agreements, all of which pose significant challenges to collaborative research initiatives. Moreover, ethical board approvals, based on detailed study protocols, are often necessary. Altogether, fulfilling all legal and ethical requirements can easily last 1 to-2 years in practice.

Subsequently, access to data often further necessitates additional agreements defining the terms of, eg, costs, split of intellectual property rights, and liability, which require additional time, effort, and trust between parties. Moreover, national variations exist in the implementation of the GDPR and the requirements for ethical board approvals throughout the EU. Overall, the current legal landscape provides robust

protection for patient data; however, its complexity complicates and slows down innovation projects, and problems multiply with the number of involved parties, sectors, and jurisdictions.

## Technical Challenges

From a technical point of view, differences in data formats and standards are further hurdles to realize the requested portability of health data across the EU. Hence, initiatives like the Observational Medical Outcomes Partnership (OMOP) aim at establishing a common data model (CDM) designed to standardize the structure and content of observational health data [5]. In consequence, multiple initiatives (eg, EHDEN [6] and DARWIN-EU [7]) have been initiated to map real-world data collected during routine health care to OMOP and to develop the corresponding harmonization processes [8]. These initiatives thus address semantic interoperability of health data as a prerequisite for portability. However, data harmonization processes are not only complex, time consuming, and error prone, but also raise the question of how to deal with valuable non-standard data elements, which may only exist at national or institutional levels. The EHDS could amplify the impact of data harmonization initiatives

through the adoption of standards and technical requirements across EU countries.

## Decentralized Data Analysis as an Enabling Technology

### Background

Decentralized data analytical approaches, including federated machine learning (FL) [9], offer a solution to overcome legal and technical challenges by enabling collaborative data analysis without necessitating centralized data storage. In FL, models are trained locally on data held by individual organizations. Only the model updates, not the raw data, are shared with a central server, which orchestrates the training process. FL enhances privacy by keeping sensitive data on-premises while enabling collaborative model training. Swarm learning (SL) [10] is a FL variant that removes the need for a central coordinating server. Instead, a blockchain is used to ensure direct and secure communication between local servers. SL offers improved scalability and faster model convergence compared to traditional FL, because it supports asynchronous learning and thus reduces the communication bottleneck with a central server. Major commercial players today provide robust libraries for federated data analysis, including descriptive summary statistics and FL/SL, thus lowering the barrier to employing such technologies in practice. Projects aiming for the implementation of EHDS-compatible platforms can thus build on relatively mature technology here.

### Concerns in Decentralized Data Analysis

Although decentralized data analysis reduces the need for data transfer, it is not immune to risks. For example, machine learning (ML) model updates can inadvertently reveal sensitive information through reverse engineering or gradient leakage [11]. Consequently, approaches such as differential privacy and encryption technologies remain essential safeguards. Differential privacy amounts to adding a defined amount of noise to model gradients during local model updates. However, this can degrade model performance and can only be mitigated by prolonged model training, which often proves computationally impractical.

Another challenge arises from the propagation of biases inherent in local datasets to the aggregated global model. This issue can be intensified when larger datasets have a disproportional influence. Furthermore, variations in data quality and heterogeneity across sites can affect ML model performance, necessitating rigorous validation and harmonization efforts across all parties participating in a decentralized data analysis. At the same time, the identification of possible reasons for biases in an ML model or generally poor prediction performance is extremely hard to detect if data scientists have no direct access to data.

While recent techniques allow the provision of human-understandable explanations of predictions produced by neural networks and can thus help to identify possibly model biases [12-14], these methods are not guaranteed to conform

between models trained on different local and pooled datasets [11]. FL/SL approaches are thus in a certain tension with the general ambition of trustworthiness of AI formulated by the EU [15].

### Technical Requirements

On a technical level, decentralized data analysis, including FL/SL, requires setting up a cloud computing infrastructure. For this purpose, multiple commercial platforms are now available. In addition, European projects such as IDERHA (Integration of Heterogeneous Data and Evidence towards Regulatory and HTA Acceptance [16]) aim for setting up a scalable cloud computing infrastructure with security measures that meet the requirements of the European health care sector. Still, there is a need to critically evaluate those solutions regarding computational infrastructure (eg, support of GPU usage), security of communication between organizations, authentication of users, and support of FL/SL. Notably, the latter may require opening specific ports at each of the participating sites, which in turn may only be possible on specific servers that are hosted in a demilitarized zone. Finally, interoperability is key. Hence, it is essential to map data at each participating organization to a standardized CDM such as OMOP.

### Legal Requirements

The legal challenges outlined above necessitate that research projects address technical and legal data governance from the outset, as they are integral components of improving access to and secondary use of health data. Effective legal facilitation of data-driven health innovation must ensure robust protection of data subject rights while mitigating compliance risks for involved entities. Simultaneously, the legal governance framework should be designed to minimize overregulation and redundancy, as such issues directly undermine the workability and impact of the technical solutions it aims to support. Given the complexity of the current regulatory and technical landscape, a scalable and modular governance approach is essential. IDERHA has developed and successfully piloted such a data governance model, which has already been adopted by several other research initiatives, such as CERTAINTY [17].

### Existing Example Projects

As of January 2025, the CORDIS web portal of the EU lists 14 projects in the health care domain after using the search terms “federated learning,” “swarm learning,” and “federated machine learning” with an OR conjunction. Following manual inspection, 8 of them mention FL in the description of their objectives. Among those 8 projects (AI-SPRINT, dAlbetes, UMBRELLA, NextGen, SEARCH, DTRIP4H, INCISIVE, and ORCHESTRA), two (INCISIVE and ORCHESTRA) have recently been completed. This shows the topicality of FL/SL concepts while at the same time highlighting that most applications of FL/SL in health care are still in an early exploratory phase.

INCISIVE [18] delivered a federated data analysis platform for artificial intelligence (AI)-based diagnosis of different cancers (breast, lung, colorectal, and prostate cancers), mostly focusing on medical images. After registration, the user can search through the available data and train AI/ML models. Moreover, organizations can contribute their own data and become a member of the cloud infrastructure.

ORCHESTRA [19] focused on the compilation and federated analysis of an EU-wide cohort to support research on SARS-CoV-2. Users can get a high-level overview about the studies included, including a standardized list of variables. Furthermore, they can apply to access specific studies.

The German Center for Neurodegenerative Diseases (DZNE) currently uses SL in the context of the early detection of Alzheimer disease, Parkinson disease, COVID-19, long COVID syndromes, leukemia, and infectious diseases [20].

Hussein et al [2] outline additional initiatives, and further results are anticipated in the coming years as ongoing projects conclude.

While this discussion mainly focuses on the connection of FL/SL with the ambitions of the EHDS, it is important to mention that FL/SL platforms are also developed in other regions of the world. For example, the Mayo clinics in the USA launched an FL platform to predict response to chemotherapy [21]. The Intel Labs and the Medical School of the University of Pennsylvania initiated a large FL study involving medical imaging data from 71 sites across six continents [22]. In addition, commercial players are now offering FL platforms.

## Challenges for Translation Into Market-Ready Solutions

The discussed examples show the first successful implementations of decentralized data analysis as proof of concept in research. Nevertheless, we expect a long way to go until market-ready high-performance infrastructures for the analysis of health data at whole population levels across Europe are available. At this point, specifically, FL/SL generates internet traffic, which increases with the number of model parameters. This can significantly slow down computation and thus negatively impact computationally intensive AI/ML applications.

Some EU countries allow the analysis of EHRs from national health registries only within authenticated environments. For instance, in Finland, health data for secondary use must be processed in secure processing environments that do not allow analysis results or model parameters to automatically leave or enter the environment. Currently, this prevents Finnish participation in a decentralized data analysis infrastructure. The EHDS aims for data access and processing within secure processing environments provided by health data access bodies. To enable the vision of the EHDS, such secure processing environments must in the future allow the use of decentralized data analysis, including FL/SL.

Further challenges arise as soon as AI/ML models should be deployed for routine health care, which is essential for generating real benefit for patients. AI/ML models used in health care must comply with the Medical Device Regulation [23], which mandates rigorous validation and monitoring processes to ensure safety and efficacy. Furthermore, the EU AI Act classifies health care-related AI/ML models as high-risk systems [24]. This imposes additional requirements for transparency, accountability, and risk management, increasing the burden on developers.

## Conclusions

The EHDS represents a bold vision for the future of health care in Europe, including control over personal health data, development of a market for EHRs, and facilitating secondary data use. These ambitions have to be balanced against concerns regarding data privacy and compliance with regulatory frameworks.

Although decentralized data analysis, including FL/SL, is not free of technical concerns and despite the limitation of trustworthiness of AI solutions, decentralized data analysis offers a promising solution to these legal challenges. However, its implementation requires significant investment in scalable technical infrastructure, data standardization, and data governance. While first example projects and commercial solutions for specific use cases and with selected partner organizations exist, the application of FL/SL in health care is generally in an early, exploratory phase. Scaling existing prototypes to enable whole population data analysis will require massive investments in technical infrastructure. In addition, deployment of AI/ML models trained in a decentralized manner for routine health care comes with additional technical, practical, and regulatory challenges.

Currently, long-lasting legal and regulatory processes for the setup of decentralized data analysis platforms cannot be circumvented, and differences in national legislation across EU member states prevent an equitable participation in decentralized data analysis initiatives. This situation can be viewed as a competitive disadvantage compared to other areas of the world. While scientific progress will likely reduce some of the technical constraints, decision makers should think about the possibilities to strengthen the predictability and coherence of the EU and national legislations stemming from the EU Digital Health and the EU Digitalization Strategies. Organizations must navigate an increasingly complex and quickly evolving legal landscape, in which they have to comply with the GDPR, the EHDS, and the AI Act in the future. While projects like IDERHA have come up with constructs for sharing data in compliance with the current legislation, the impact of upcoming legislations remains to be determined. Altogether, there is a strong need for a better harmonization, stabilization, and simplification of the EU legislation to make the EHDS a success story. In this regard, the decentralized analysis of health data across EU countries has to be legally enabled.



In conclusion, there is a long way to go before the EHDS can deliver on its promise of a better connected and efficient health ecosystem across Europe.

## Acknowledgments

This project has received funding from the European Union's Horizon Europe research and innovation program under grant agreement No [101112135] (Integration of heterogeneous Data and Evidence towards Regulatory and HTA Acceptance [IDERHA]) through the Innovative Health Initiative (IHI) Joint Undertaking (JU). Support is also received from life science industries represented by COCIR, EFPIA/Vaccines Europe, EuropaBio, and MedTech Europe. Support is also received from our Swiss and UK partners.

The writing of this text was assisted by GPT-4o (OpenAI, 2025).

## Disclaimer

This paper reflects the personal opinion of the authors but not necessarily of their employers.

## Conflicts of Interest

GJ is an employee of Novartis, and SN and CM are employees of Johnson & Johnson Medical GmbH. The paper expresses the personal opinion of the authors but not of the companies.

## References

1. European Health Data Space Regulation (EHDS). European Commission. URL: [https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space\\_en](https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en) [Accessed 2025-01-17]
2. Hussein R, Balaur I, Burmann A, et al. Getting ready for the European Health Data Space (EHDS): IDERHA's plan to align with the latest EHDS requirements for the secondary use of health data. Open Research Europe. URL: <https://open-research-europe.ec.europa.eu/articles/4-160> [Accessed 2025-02-10]
3. Second joint action towards the European Health Data Space – TEHDAS2. Tehdas. URL: <https://tehdas.eu/> [Accessed 2025-03-03]
4. 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) (Text with EEA Relevance). Vol 119. European Union; 2016. URL: <http://data.europa.eu/eli/reg/2016/679/oj/eng> [Accessed 2018-02-06]
5. OMOP common data model. URL: <https://ohdsi.github.io/CommonDataModel/index.html> [Accessed 2025-01-20]
6. European Health Data Evidence Network. URL: <https://www.ehden.eu/> [Accessed 2025-03-03]
7. The DARWIN EU® data network. DARWIN EU. URL: <https://darwin-eu.org/index.php/data/data-network> [Accessed 2024-08-08]
8. Henke E, Zoch M, Peng Y, Reinecke I, Sedlmayr M, Bathelt F. Conceptual design of a generic data harmonization process for OMOP common data model. BMC Med Inform Decis Mak. Feb 26, 2024;24(1):58. [doi: [10.1186/s12911-024-02458-7](https://doi.org/10.1186/s12911-024-02458-7)] [Medline: [38408983](https://pubmed.ncbi.nlm.nih.gov/38408983/)]
9. Prayitno, Shyu CR, Putra KT, et al. A systematic review of federated learning in the healthcare area: from the perspective of data properties and applications. Appl Sci (Basel). Jan 2021;11(23):11191. [doi: [10.3390/app112311191](https://doi.org/10.3390/app112311191)]
10. Warnat-Herresthal S, Schultze H, Shastry KL, et al. Swarm learning for decentralized and confidential clinical machine learning. Nature New Biol. Jun 2021;594(7862):265-270. [doi: [10.1038/s41586-021-03583-3](https://doi.org/10.1038/s41586-021-03583-3)] [Medline: [34040261](https://pubmed.ncbi.nlm.nih.gov/34040261/)]
11. Bak M, Madai VI, Celi LA, et al. Federated learning is not a cure-all for data ethics. Nat Mach Intell. Apr 2024;6(4):370-372. [doi: [10.1038/s42256-024-00813-x](https://doi.org/10.1038/s42256-024-00813-x)]
12. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, editors. Adv Neural Inf Process Syst. Curran Associates, Inc; 2017:4765-4774. URL: <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf> [Accessed 2025-06-24]
13. Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. arXiv. Preprint posted online on Jun 13, 2017. [doi: [10.48550/arXiv.1703.01365](https://doi.org/10.48550/arXiv.1703.01365)]
14. Binder A, Montavon G, Bach S, Müller KR, Samek W. Layer-wise relevance propagation for neural networks with local renormalization layers. arXiv. Preprint posted online on Apr 4, 2016. [doi: [10.48550/arXiv.1604.00825](https://doi.org/10.48550/arXiv.1604.00825)]
15. Ethics guidelines for trustworthy AI. European Commission. 2023. URL: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> [Accessed 2023-08-31]
16. Welcome to IDERHA. IDERHA. URL: <https://www.iderha.org/integrating-health-data> [Accessed 2025-01-20]
17. CERTAINTY | Virtual twin for personalised cancer treatment. CERTAINTY. URL: <https://www.certainty-virtualtwin.eu> [Accessed 2025-05-23]

18. Incisive Project. URL: <https://incisive-project.eu/> [Accessed 2025-01-24]
19. ORCHESTRA - EU horizon 2020 cohort to tackle COVID-19 internationally. ORCHESTRA. URL: <https://orchestra-cohort.eu/> [Accessed 2025-05-16]
20. DZNE swarm learning hub. DZNE. URL: <https://www.dzne.de/en/swarm-learning-hub/welcome/> [Accessed 2025-05-16]
21. Anastasijevic D. Mayo Clinic launches its first platform initiative. Mayo Clinic News Network. Jan 14, 2020. URL: <https://newsnetwork.mayoclinic.org/discussion/mayo-clinic-launches-its-first-platform-initiative/> [Accessed 2025-05-16]
22. Pati S, Baid U, Edwards B, et al. Federated learning enables big data for rare cancer boundary detection. Nat Commun. Dec 5, 2022;13(1):7346. [doi: [10.1038/s41467-022-33407-5](https://doi.org/10.1038/s41467-022-33407-5)] [Medline: [36470898](https://pubmed.ncbi.nlm.nih.gov/36470898/)]
23. Regulation - 2017/745 - EN - medical device regulation. EUR-Lex. URL: <https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng> [Accessed 2025-01-24]
24. Reflection paper on the use of artificial intelligence (AI) in the medicinal product lifecycle (draft). European Medicines Agency. 2023. URL: <https://www.ema.europa.eu/en/news/reflection-paper-use-artificial-intelligence-lifecycle-medicines> [Accessed 2025-06-23]

## Abbreviations

**AI:** artificial intelligence  
**CDM:** common data model  
**DZNE:** German Center for Neurodegenerative Diseases  
**EHDS:** European Health Data Space  
**EHR:** electronic health record  
**EU:** European Union  
**FL:** federated machine learning  
**GDPR:** General Data Protection Regulation  
**IDERHA:** Integration of Heterogeneous Data and Evidence towards Regulatory and HTA Acceptance  
**ML:** machine learning  
**OMOP:** Observational Medical Outcomes Partnership  
**SL:** swarm learning

*Edited by Amaryllis Mavragani; peer-reviewed by Ankit Gupta, Sadhasivam Mohanadas, Tope Amusa; submitted 24.04.2025; final revised version received 23.05.2025; accepted 29.05.2025; published 19.09.2025*

### *Please cite as:*

*Fröhlich H, Funck Hansen A, Hilvo M, Jansen G, Madan S, Moazemi S, Negreanu S, Satagopam V, Gribbon P, Muehlendyck C*

*Reality Check: The Aspirations of the European Health Data Space Amidst Challenges in Decentralized Data Analysis*  
*J Med Internet Res 2025;27:e76491*

URL: <https://www.jmir.org/2025/1/e76491>

doi: [10.2196/76491](https://doi.org/10.2196/76491)

© Holger Fröhlich, Anne Funck Hansen, Mika Hilvo, Gunther Jansen, Sumit Madan, Sobhan Moazemi, Sanziana Negreanu, Venkata Satagopam, Phil Gribbon, Christian Muehlendyck. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 19.09.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org>, as well as this copyright and license information must be included.