FL-EHDS: A Privacy-Preserving Federated Learning Framework for the European Health Data Space

Fabio Liberti
Department of Computer Science
Universitas Mercatorum, Rome, Italy
fabio.liberti@unimercatorum.it
ORCID: 0000-0000-0000-0000

***Abstract*—The European Health Data Space (EHDS), effective March 2025, mandates cross-border health data analytics while preserving privacy. Federated Learning (FL) emerges as the key enabling technology, yet a systematic evidence synthesis reveals critical gaps: only 23% of FL implementations achieve production deployment, with hardware heterogeneity (78%) and non-IID data (67%) as dominant barriers. Legal uncertainties regarding gradient data status under GDPR remain unresolved. We present FL-EHDS, a three-layer compliance framework integrating governance mechanisms (HDABs, data permits), FL orchestration (aggregation within Secure Processing Environments), and data holder components. The framework maps evidence-based barriers to specific mitigation strategies and provides compliance checkpoints aligned with EHDS requirements. We contribute: (1) first systematic barrier taxonomy for FL in EHDS contexts; (2) a reference architecture addressing identified gaps; (3) an implementation roadmap for the 2025-2031 transition period. FL-EHDS bridges the technology-governance divide critical for successful EHDS operationalization.**

***Index Terms*—Federated Learning, European Health Data Space, Privacy-Preserving Technologies, GDPR Compliance, Health Data Governance, Cross-Border Analytics**

I. Introduction

A. The EHDS Challenge

The European Health Data Space (EHDS), established by Regulation (EU) 2025/327 and entering into force on 26 March 2025, represents the EU's most ambitious initiative for cross-border health data governance [1]. The regulation creates a dual framework: primary use through MyHealth@EU for patient care coordination, and secondary use through HealthData@EU for research, innovation, and public health surveillance.

This ambition confronts a fundamental tension: maximizing data utility while preserving privacy and maintaining data sovereignty across 27 Member States with diverse health systems and regulatory traditions.

B. The Technology-Governance Divide

Federated Learning (FL) emerges as the theoretically ideal solution—the model travels to data rather than centralizing sensitive records [2]. However, policy assumptions about technological maturity may not align with implementation realities.

Recent evidence reveals a sobering gap: only 23% of FL implementations achieve sustained production deployment in healthcare settings [3]. Technical barriers including hardware heterogeneity (affecting 78% of pilot participants) and non-IID data challenges (67%) persist [3]. Simultaneously, legal uncertainties regarding gradient data status under GDPR remain unresolved [4].

The scholarly discourse is characterized by disciplinary fragmentation: technical literature develops sophisticated algorithms while abstracting from legal constraints; legal scholarship analyzes compliance without engaging technical feasibility. This divide risks producing recommendations that are either technically naive or legally ungrounded.

C. Contributions

This paper makes three primary contributions:

1) **Systematic Barrier Taxonomy**: First evidence synthesis of FL implementation barriers specific to EHDS contexts, based on systematic review of 47 documents (PRISMA methodology, GRADE-CERQual confidence assessment).
2) **FL-EHDS Framework**: A three-layer reference architecture mapping identified barriers to specific mitigation strategies, with compliance checkpoints aligned with GDPR and EHDS requirements.
3) **Implementation Roadmap**: Prioritized actions for the 2025-2031 transition period, with stakeholder-specific recommendations and evaluation metrics.

The remainder of this paper is organized as follows: Section II provides essential background on EHDS and FL fundamentals; Section III presents the FL-EHDS framework architecture; Section IV synthesizes barrier evidence; Section V outlines the implementation roadmap; Section VI discusses implications and limitations.

II. Background

A. European Health Data Space

The EHDS establishes Health Data Access Bodies (HDABs) in each Member State to evaluate and authorize secondary use requests through standardized data permits. Article 53 enumerates permitted purposes including scientific research, public health surveillance, and AI training for medical devices. Article 71 introduces an opt-out mechanism balancing collective benefit with individual autonomy [5].

Table I summarizes the implementation timeline creating urgency for FL readiness.

TABLE I
EHDS IMPLEMENTATION TIMELINE

Date	Milestone	FL Relevance
Mar 2025	Entry into force	Framework active
Mar 2027	Delegated acts	Legal clarification
Mar 2029	Secondary use	FL operational
Mar 2031	Sensitive data	Genetic, imaging

B. Federated Learning Fundamentals

FL inverts the traditional paradigm: rather than centralizing data, the model travels to distributed sources [6]. Local training produces gradients capturing learned patterns without raw records; these are aggregated centrally and redistributed for iterative refinement [7].

Several characteristics render FL suitable for EHDS: alignment with GDPR data minimization (Article 5(1)(c)); scalability to pan-European scope; and applicability to EHDS permitted purposes. However, known challenges include non-IID data distributions [8], communication overhead, and privacy attacks (gradient inversion [9], membership inference [10]).

III. FL-EHDS FRAMEWORK

We present FL-EHDS, a three-layer compliance framework designed specifically for EHDS cross-border analytics. The architecture integrates governance mechanisms, FL orchestration, and data holder components with compliance-by-design principles.

A. Architecture Overview

?? illustrates the FL-EHDS architecture. The framework comprises:

- **Layer 1 (Governance)**: HDAB integration, data permit workflows, opt-out registry
- **Layer 2 (Orchestration)**: Aggregation within Secure Processing Environments
- **Layer 3 (Data Holders)**: Local training engines, FHIR preprocessing

B. Layer 1: Governance Layer

The Governance Layer provides the regulatory interface between FL operations and EHDS requirements:

HDAB Integration: Standardized APIs enable automated data permit verification before training initiation. Multi-HDAB synchronization protocols coordinate cross-border studies involving multiple Member States.

Opt-out Registry: The framework consults national opt-out registries before including data from any individual, ensuring Article 71 compliance. Technical mechanisms handle granularity (purpose-specific vs. blanket opt-out) and temporal dynamics (participants opting out mid-study).

Data Permit Workflow: Authorization requests are routed through appropriate HDABs with purpose limitation enforcement throughout the FL lifecycle.

C. Layer 2: FL Orchestration Layer

The Orchestration Layer operates within Secure Processing Environments (SPEs) as mandated by EHDS:

Aggregation Module: Implements FedAvg [6] with FedProx [8] adaptations for non-IID data. Gradient compression reduces communication overhead for cross-border transmission.

Privacy Protection: Differential privacy integration with configurable ε-budget management; gradient clipping limits information leakage; membership inference defense validates model privacy before release.

Compliance Modules: Audit logging aligned with GDPR Article 30; purpose limitation enforcement throughout training; automated compliance verification at each aggregation round.

D. Layer 3: Data Holder Layer

The Data Holder Layer addresses infrastructure heterogeneity:

Adaptive Training Engine: Resource-aware model partitioning accommodates hardware heterogeneity—from GPU clusters at academic centers to CPU-only environments in community hospitals.

FHIR-Native Preprocessing: Data normalization to HL7 FHIR ensures semantic interoperability; quality assessment validates training-readiness.

Secure Communication: Encrypted gradient transmission with integrity verification; no raw data leaves institutional boundaries.

IV. EVIDENCE SYNTHESIS: BARRIERS AND MITIGATION

A. Methodology Summary

We conducted a systematic review following PRISMA 2020 guidelines, searching eight databases (PubMed, IEEE Xplore, Scopus, Web of Science, ScienceDirect, Springer Nature, Frontiers, arXiv) for publications from May 2022 to January 2026. From 847 identified records, 47 documents met inclusion criteria (44 peer-reviewed). Quality was assessed using MMAT; confidence in findings using GRADE-CERQual.

B. Technical Barriers

Table II summarizes the barrier taxonomy with evidence and framework mitigation.

TABLE II
FL IMPLEMENTATION BARRIERS FOR EHDS

Barrier	Prev.	Evidence	Mitigation
Hardware heterogeneity	78%	Fröhlich 2025	Adaptive engine
Non-IID data	67%	Multiple	FedProx
Production gap	23%	Fröhlich 2025	Ref. impl.
FHIR compliance	34%	Hussein 2025	Preprocessing

The 23% production deployment rate [3] indicates most FL systems remain at pilot stages, with direct implications for the March 2029 deadline.

C. Legal Uncertainties

Three legal questions remain unresolved with moderate GRADE-CERQual confidence:

Gradient Data Status: Whether gradients constitute personal data under GDPR. Research demonstrates reidentification risks through gradient inversion [9] and membership inference [10], yet no DPA has issued formal guidance [4].

Model Anonymity: Whether aggregated models qualify as anonymous or pseudonymous, determining GDPR applicability to model sharing.

Controller Responsibilities: Unclear allocation of controller/processor roles in multi-party FL configurations.

The FL-EHDS framework addresses these through explicit compliance checkpoints and audit trails, pending regulatory clarification in 2027 delegated acts.

D. Organizational Barriers

HDAB capacity asymmetries documented by TEHDAS [11] show Nordic countries 2-3 years ahead of Southern/Eastern European states. Access timelines vary from 3 weeks (Finland) to over 12 months (France) [12]—organizational factors, not technical limitations, dominate.

V. Implementation Roadmap

A. Phased Implementation

Table III presents a phased implementation strategy aligned with EHDS milestones.

TABLE III
FL-EHDS Implementation Roadmap

Phase	Timeline	Actions
Foundation	2025-26	Reference implementation; pilot deployment
Clarification	2027	Delegated acts integration; legal guidance
Scaling	2028-29	Multi-MS deployment; HDAB onboarding
Operation	2029-31	Full production; all data categories

B. Stakeholder Recommendations

EU Policymakers: Clarify gradient data status in 2027 delegated acts; establish FL-specific HDAB guidance; fund cross-border pilot programs with mandatory public reporting.

National Authorities: Invest early in HDAB capacity; train staff on FL technical evaluation; standardize authorization workflows.

Healthcare Organizations: Accelerate FHIR compliance; assess local training infrastructure; participate in pilot programs.

Researchers: Shift from proof-of-concept to implementation studies; report negative results; pursue interdisciplinary collaboration.

C. Evaluation Metrics

Key success metrics include: production deployment rate (target: ¿50% by 2029, vs. 23% baseline); FHIR compliance (¿70% vs. 34%); legal clarity (resolved vs. unresolved); cross-border pilots (20+ use cases vs. 5 current).

VI. Discussion and Conclusions

A. Key Findings

This paper contributes the first systematic synthesis of FL barriers specific to EHDS implementation, revealing that **legal uncertainties—not technical barriers—are the critical blocker**. The 23% production deployment rate indicates technology readiness issues, but the more fundamental challenge lies in unresolved regulatory questions about gradient data protection status.

The technology-governance divide poses risks for both effectiveness and legitimacy. FL cannot substitute for robust governance; public trust depends on institutional transparency rather than technical privacy guarantees alone [13].

B. Limitations

The FL-EHDS framework has not yet been empirically validated in operational EHDS environments—such validation must await the 2029 application date. The evidence base is predominantly anticipatory, analyzing the newly-adopted regulation rather than actual implementation. The rapidly evolving technical and regulatory landscape may outpace our findings.

C. Future Work

Future work should prioritize: empirical validation through HealthData@EU pilot integration; citizen attitude studies examining FL acceptance; economic sustainability modeling for HDAB FL operations; and longitudinal tracking as implementation proceeds through 2031.

D. Conclusion

The FL-EHDS framework bridges the technology-governance divide by providing a reference architecture with explicit compliance checkpoints. The 2027 delegated acts represent a critical window for legal clarification; without it, the 2029 deadline arrives with FL adoption still inhibited by uncertainty.

For the FL research community, the pressing question is no longer "can FL work?" but "how does FL work in practice under real regulatory constraints?" This paper provides an evidence-based foundation for that investigation.

Acknowledgments

The author thanks the TEHDAS Joint Action consortium for making preparatory materials publicly available.

References

[1] European Commission, “Regulation (EU) 2025/327 on the European Health Data Space,” *Official Journal of the European Union*, 2025.

[2] N. Rieke et al., “The future of digital health with federated learning,” *NPJ Digital Medicine*, vol. 3, article 119, 2020.

[3] H. Fröhlich et al., “Reality check: The aspirations of the EHDS amidst challenges in decentralized data analysis,” *J. Medical Internet Research*, vol. 27, e76491, 2025.

[4] P. Quinn et al., “Will the GDPR restrain health data access bodies under the EHDS?” *Computer Law & Security Review*, vol. 54, 105993, 2024.

[5] C. Staunton et al., “Ethical and social reflections on the proposed European Health Data Space,” *Eur. J. Human Genetics*, vol. 32, no. 5, pp. 498–505, 2024.

[6] B. McMahan et al., “Communication-efficient learning of deep networks from decentralized data,” *Proc. AISTATS*, pp. 1273–1282, 2017.

[7] P. Kairouz et al., “Advances and open problems in federated learning,” *Foundations and Trends in ML*, vol. 14, no. 1–2, pp. 1–210, 2021.

[8] T. Li et al., “Federated optimization in heterogeneous networks,” *Proc. MLSys*, vol. 2, pp. 429–450, 2020.

[9] L. Zhu et al., “Deep leakage from gradients,” *Advances in NeurIPS*, vol. 32, pp. 14774–14784, 2019.

[10] R. Shokri et al., “Membership inference attacks against machine learning models,” *IEEE S&P*, pp. 3–18, 2017.

[11] TEHDAS Joint Action, “Are EU member states ready for the European Health Data Space?” *Eur. J. Public Health*, vol. 34, no. 6, pp. 1102–1108, 2024.

[12] R. B. Forster et al., “User journeys in cross-European secondary use of health data,” *Eur. J. Public Health*, vol. 35, Suppl. 3, pp. iii18–iii24, 2025.

[13] S. van Drumpt et al., “Secondary use under the European Health Data Space: Setting the scene and towards a research agenda on privacy-enhancing technologies,” *Frontiers in Digital Health*, vol. 7, 1602101, 2025.

[14] R. Hussein et al., “Interoperability framework of the EHDS for secondary use,” *J. Medical Internet Research*, vol. 27, e69813, 2025.