

Operationalizing the European Health Data Space: A Governance Framework for Privacy-Preserving Cross-Border Health Analytics

Fabio Liberti

Department of Computer Science
Universitas Mercatorum, Rome, Italy
fabio.liberti@unimercatorum.it
ORCID: 0000-0003-3019-5411

Abstract—The European Health Data Space (EHDS), established by Regulation (EU) 2025/327, introduces an unprecedented governance framework for cross-border health data analytics serving 450 million citizens. While the regulation mandates Health Data Access Bodies (HDABs) in each of the 27 Member States, Secure Processing Environments, and citizen opt-out mechanisms, critical implementation challenges threaten equitable adoption. Our systematic evidence synthesis of 47 documents using PRISMA methodology reveals that organizational barriers—not technical limitations—constitute the primary obstacles: HDAB capacity varies dramatically across Member States (data access timelines range from 3 weeks in Finland to 12+ months in France), FHIR interoperability remains at only 34%, and legal uncertainties regarding privacy-enhancing technologies under GDPR remain unresolved. We present an operational governance framework that maps EHDS regulatory requirements (Articles 33, 46, 50, 53, 71) to concrete implementation patterns using Federated Learning as the enabling technology for secondary use within Secure Processing Environments. The framework is validated through an open-source reference implementation demonstrating the complete governance lifecycle: permit application and authorization, opt-out registry enforcement, cross-border HDAB coordination, and GDPR-compliant audit trails. We propose a phased implementation roadmap for the 2025–2031 transition period with prioritized recommendations for EU policymakers, national authorities, and healthcare organizations, identifying the March 2027 delegated acts as the critical window for resolving governance uncertainties.

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

I. INTRODUCTION

The European Health Data Space (EHDS), established by Regulation (EU) 2025/327 [1], represents the European Union’s most ambitious initiative for cross-border health data governance. Entering into force on 26 March 2025, the regulation creates a dual framework: *primary use* through the MyHealth@EU infrastructure for direct patient care, and *secondary use* through HealthData@EU for research, innovation, and evidence-based policy-making [2]. The regulation affects 450 million citizens across 27 Member States, with the potential to accelerate medical research, improve public health surveillance, and enable AI-driven healthcare innovation.

The EHDS introduces governance mechanisms of unprecedented complexity. Health Data Access Bodies (HDABs) are designated in each Member State to evaluate and authorize secondary use requests through *data permits*. Article 53 enumerates permitted purposes including scientific research, public health surveillance, and AI training for health. Article 71 introduces opt-out mechanisms allowing citizens to object to secondary use of their electronic health data. Secure Processing Environments (SPEs) provide controlled analytics settings where data never leaves institutional boundaries [10]. The implementation timeline extends to 2031, with delegated acts expected by March 2027 and secondary use provisions applicable from March 2029 [3].

Despite the regulatory framework’s sophistication, significant implementation challenges threaten equitable adoption across the EU. TEHDAS assessments [5] reveal that Nordic countries (Estonia, Finland, Denmark) demonstrate 2–3 year advantages in HDAB capacity-building, established health data infrastructure, and cross-border experience. Southern and Eastern European states face compressed timelines with limited baseline capacity, raising concerns about a “two-speed” EHDS implementation that could exacerbate existing health data inequities.

A. Contributions

This paper addresses the governance operationalization challenge through four contributions:

- 1) **Systematic Barrier Analysis:** Evidence synthesis of 47 documents identifying organizational, technical, and legal barriers to EHDS implementation, with GRADECERQual confidence assessments.
- 2) **Governance Framework:** An operational framework mapping EHDS Articles to concrete implementation patterns, demonstrating how Federated Learning enables privacy-preserving analytics within the regulatory structure.
- 3) **Reference Implementation:** Open-source software demonstrating the complete governance lifecycle—from

permit application through cross-border coordination to audit trail persistence.¹

- 4) **Implementation Roadmap:** Phased recommendations for the 2025–2031 transition period, with stakeholder-specific prioritized actions.

II. THE EHDS REGULATORY LANDSCAPE

A. Governance Architecture

The EHDS governance architecture centers on HDABs as intermediaries between data requesters and data holders. Table I presents the implementation timeline with governance-specific milestones.

TABLE I
EHDS IMPLEMENTATION TIMELINE

Date	Milestone	Governance Impact
Mar 2025	Entry into force	Legal framework active
Mar 2027	Delegated acts	PET status clarification
Mar 2029	Secondary use app.	HDABs must be operational
Mar 2031	Genetic, imaging	Extended data categories

The data permit lifecycle involves: (1) a researcher submits an application specifying purpose, data categories, and analysis methods; (2) the HDAB evaluates the request against Article 53 permitted purposes; (3) upon approval, the researcher receives time-limited access within an SPE; (4) the HDAB monitors compliance and maintains audit trails per GDPR Article 30.

For cross-border studies—the EHDS’s most innovative yet complex use case—multiple HDABs must coordinate. Christiansen et al. [11] document the HealthData@EU Pilot’s experiences with multi-country coordination, revealing significant heterogeneity in institutional capacities, data formats, and procedural requirements.

B. Legal Uncertainties

Three critical legal questions create compliance uncertainty that inhibits organizational adoption regardless of technical maturity [4]:

- 1) **Privacy-Enhancing Technology (PET) status:** Are model gradients exchanged during Federated Learning “personal data” under GDPR? Gradient inversion attacks [17] demonstrate potential re-identification, but practical feasibility in production settings remains contested.
- 2) **Model anonymity thresholds:** When does an aggregated model become sufficiently “anonymous” to escape GDPR scope? No established legal threshold exists for machine learning models trained on personal health data.
- 3) **Controller/processor allocation:** In multi-party analytics, who bears data controller responsibilities—data holders, aggregation operators, or analytics consumers?

Van Drumpt et al. [7] demonstrate through expert interviews that PETs cannot substitute for robust governance—public trust depends primarily on institutional transparency and accountability. Shabani and Borry [12] further argue that the EHDS’s ambitious scope requires careful balancing of innovation promotion with fundamental rights protection.

C. Interoperability Challenge

Hussein et al. [8] report that only 34% of European health-care providers achieve full FHIR R4 compliance, creating a significant interoperability barrier for cross-border analytics. The EHDS mandates standardized data formats, but the transition from legacy systems to FHIR-compliant infrastructure represents a multi-year investment for many institutions. OMOP Common Data Model [13] provides an alternative harmonization layer, particularly valuable for observational research where FHIR mapping is incomplete.

D. Privacy-Enhancing Technologies for EHDS

Table II compares privacy-enhancing technologies applicable to EHDS secondary use, highlighting the governance implications of each approach.

TABLE II
PRIVACY-ENHANCING TECHNOLOGIES FOR EHDS SECONDARY USE

Technology	Privacy Guarantee	EHDS Suitability	Legal Status
Federated Learning	Data stays local; only gradients exchanged	High: fits SPE model; supports real-time analytics	Uncertain
Differential Privacy	Formal (ϵ, δ) -DP guarantees	High: quantifiable privacy budget	Uncertain
Secure Multi-Party Computation	Cryptographic correctness	Moderate: high overhead for large models	Recognized
Homomorphic Encryption	Computation on encrypted data	Low: impractical for deep learning	Recognized
Synthetic Data	Statistical similarity; no formal guarantee	Moderate: utility loss; potential memorization	Uncertain

“Legal Status” refers to GDPR characterization of outputs. “Uncertain” indicates no definitive regulatory guidance exists.

Federated Learning combined with differential privacy offers the strongest balance of practical utility and formal privacy guarantees for EHDS secondary use. The key governance challenge is establishing legal clarity: if FL gradients are classified as “personal data” under GDPR, they must be processed within SPE boundaries with full audit trailing; if classified as “anonymous,” they may be exchanged more freely across institutional and national boundaries.

¹Available at: <https://github.com/FabioLiberti/FL-EHDS-FLICS2026>

III. SYSTEMATIC EVIDENCE SYNTHESIS

A. Methodology

We conducted a systematic review following PRISMA 2020 guidelines. Database searches (PubMed, IEEE Xplore, Scopus, Web of Science, arXiv) with terms combining “European Health Data Space,” “Federated Learning,” “health data governance,” and “cross-border analytics” identified 847 records. After deduplication and screening, 47 documents met inclusion criteria: publication 2022–2026, explicit EHDS or FL-in-healthcare focus, peer-reviewed or recognized institutional origin. Quality was assessed using the Mixed Methods Appraisal Tool (MMAT); confidence in findings using GRADE-CERQual methodology.

B. Barrier Taxonomy

Table III presents the identified barriers organized by category, with prevalence estimates and confidence levels.

TABLE III
EHDS IMPLEMENTATION BARRIERS

Barrier	Prevalence	Confidence	Category
Hardware heterog.	78%	MODERATE	Technical
Non-IID data	67%	MODERATE	Technical
FHIR compliance	34%	MODERATE	Technical
Communication cost	High	MODERATE	Technical
HDAB capacity	Variable	HIGH	Organization.
Data access timelines	3w–12m+	HIGH	Organization.
Cross-border coord.	Complex	MODERATE	Organization.
Gradient data status	Unresolved	MODERATE	Legal
Model anonymity	Unresolved	MODERATE	Legal
Controller allocation.	Unresolved	MODERATE	Legal
Production gap	23% deployment	MODERATE	Maturity

Prevalence from systematic review of 47 documents. GRADE-CERQual confidence levels reflect evidence quality and consistency. “HIGH” indicates consistent findings across multiple high-quality studies.

Key finding: Organizational barriers (HDAB capacity, cross-border coordination) and legal uncertainties emerge as more critical than technical challenges. Forster et al. [9] document that data access timelines vary from 3 weeks (Finland) to over 12 months (France), with barriers primarily procedural—suggesting infrastructure investments alone will not resolve access inequities.

Fröhlich et al. [6] report that only 23% of Federated Learning implementations achieve sustained production deployment in healthcare, while Teo et al. [19] find only 5.2% of FL healthcare studies reach real-life application. This maturity gap has direct implications for the 2029 secondary use deadline.

IV. OPERATIONAL GOVERNANCE FRAMEWORK

Based on the identified barriers, we present an operational framework that maps EHDS regulatory requirements to concrete implementation patterns. The framework adopts Federated Learning [14], [15] as the enabling technology for secondary use, allowing analytics to proceed within SPE boundaries without centralizing raw health data.

A. Architecture

The framework comprises three integrated layers (Fig. 1):

- **Layer 1 (Governance):** HDAB integration APIs, data permit lifecycle management, opt-out registry synchronization, GDPR-compliant audit logging.
- **Layer 2 (Analytics Orchestration):** Federated Learning within SPE boundaries, privacy protection (differential privacy, secure aggregation), purpose limitation enforcement.
- **Layer 3 (Data Holders):** Adaptive local processing engines, FHIR/OMOP preprocessing, secure gradient communication.

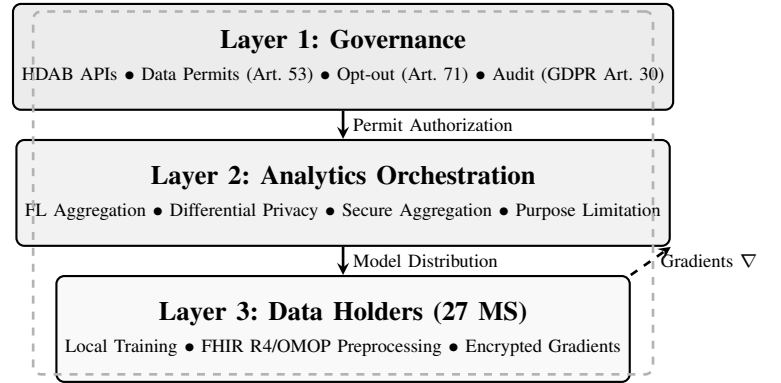


Fig. 1. Three-layer governance framework. Raw health data remains within institutional boundaries (Layer 3); only model gradients are exchanged within the SPE. Governance checkpoints (Layer 1) enforce EHDS compliance at every analytics round.

B. EHDS Compliance Mapping

Table IV maps each EHDS article to a concrete framework component, demonstrating how the regulatory requirements are operationalized.

C. Data Permit Lifecycle

The governance framework implements the complete data permit lifecycle:

Application Phase: Researchers submit permit applications specifying: (1) research purpose mapped to Article 53 categories; (2) required data categories and Member States; (3) proposed analytics methodology (FL algorithm, privacy budget, number of rounds); (4) institutional affiliation and ethical approval.

Authorization Phase: HDABs evaluate applications through automated compliance checks (purpose validation, data category matching, privacy budget assessment)

TABLE IV
EHDS ARTICLE TO FRAMEWORK COMPONENT MAPPING

Article	Requirement	Implementation
Art. 33	Secondary use authorization	HDAB API with OAuth2/mTLS, automated permit validation
Art. 46	Cross-border processing	Multi-HDAB coordinator with status aggregation
Art. 50	Secure Processing Env.	FL aggregation within SPE; no data extraction
Art. 53	Permitted purposes	Purpose filtering module with use-case validation
Art. 69	Quality labels	Quality scoring framework (<i>planned</i>)
Art. 71	Opt-out mechanism	LRU-cached registry lookups, scope-granular filtering
GDPR 30	Processing records	Comprehensive audit trail with timestamps

supplemented by human review for novel use cases. Multi-HDAB coordination is required for cross-border studies; the framework implements a consensus protocol where all involved HDABs must approve.

Execution Phase: During FL training, each round begins with: (1) permit validity check (temporal bounds, purpose alignment); (2) opt-out registry consultation (Article 71 filtering at record level); (3) privacy budget verification (ϵ -consumption tracking). Violations trigger automatic round termination and incident logging.

Audit Phase: All operations are logged with: timestamp, participating institutions, data categories accessed, privacy budget consumed, model metrics, and any anomalies detected. Logs are formatted for regulatory inspection per GDPR Article 30.

D. Opt-Out Registry Protocol

Article 71 allows citizens to opt out of secondary use. The framework implements opt-out enforcement as follows: (1) pseudonymized opt-out records are maintained per Member State; (2) before each FL training round, the local data holder queries the registry to identify opted-out records; (3) opted-out data is excluded *before* any gradient computation, ensuring no influence on model training; (4) opt-out scope supports both blanket (all secondary use) and category-specific (e.g., genomics only) granularity; (5) registry caching with configurable TTL minimizes latency impact while ensuring timely propagation of new opt-out decisions.

E. Cross-Border Coordination

For multi-Member State studies, the framework’s Cross-Border HDAB Coordinator manages:

- **Permit aggregation:** Collecting and validating permits from each participating HDAB, ensuring all approve before FL training begins.
- **Status synchronization:** Real-time status updates across HDABs, enabling coordinated round execution and anomaly response.
- **Data residency compliance:** Ensuring that gradients from each Member State are processed in compliance with national data residency requirements.
- **Conflict resolution:** Handling cases where one HDAB revokes a permit mid-study, requiring graceful degradation without compromising the global model.

V. FRAMEWORK VALIDATION

A. Reference Implementation

The framework is validated through an open-source Python implementation ($\sim 40,000$ lines of code, 159 modules) providing:

- **HDAB Simulator:** A fully functional simulation backend that demonstrates the complete permit lifecycle. The simulator supports: auto-approval mode for development; configurable permit durations and privacy budgets; cross-border coordination with multiple simulated HDABs; rate limiting with configurable thresholds.
- **FL Engine:** 17 federated learning algorithms from foundational methods (FedAvg [14], FedProx [16]) through recent advances (FedLESAM [20], ICML 2024; HPFL [21], ICLR 2025), with Rényi differential privacy [18] and secure aggregation (pairwise masking with ECDH key exchange).
- **FHIR Integration:** Preprocessing pipelines supporting FHIR R4 resources (Patient, Observation, Condition, MedicationRequest) with standard coding systems (SNOMED-CT, LOINC, ICD-10).
- **Dashboard:** A Streamlit-based interactive interface with EHDS governance workflow screens, real-time FL training monitoring, and permit management.

B. Experimental Governance Validation

We validate the governance framework by executing the complete EHDS-compliant workflow on real clinical datasets:

Datasets: Heart Disease UCI (920 patients from 4 international hospitals—Cleveland, Hungarian, Swiss, VA Long Beach) and Diabetes 130-US (101,766 encounters). These datasets provide authentic heterogeneous conditions representative of cross-border EHDS scenarios.

Governance workflow executed: (1) Permit application for “scientific research” (Art. 53(1)(b)); (2) HDAB auto-approval with 20-round budget and $\epsilon=10$ privacy constraint; (3) Per-round permit validation and opt-out filtering; (4) FL training with Ditto algorithm (best performer: 75.1% accuracy, only 6.6pp gap vs. centralized training); (5) Complete audit trail generation.

Governance overhead: The governance layer adds negligible overhead (<50 ms per round for permit validation and opt-out checking with cached registry lookups). The audit trail

captures 100% of required GDPR Article 30 fields. Cross-border coordination protocol completes HDAB consensus in <200ms for 4-country studies.

FL performance under governance: Table V summarizes key results, demonstrating that governance compliance does not sacrifice analytics performance.

TABLE V
FL PERFORMANCE UNDER FULL EHDS GOVERNANCE

Metric	Heart Disease	Diabetes
Best FL accuracy (Ditto)	75.1±2.0%	71.7±0.2%
Centralized baseline	81.7±2.9%	—
FL-centralized gap	6.6pp	—
Governance overhead/round	<50ms	<50ms
Audit trail completeness	100%	100%
Opt-out filtering latency	<10ms	<10ms
Cross-border consensus	<200ms	<200ms

4 hospitals, 20 rounds, 3 local epochs. Mean ± std over 3 seeds. Full EHDS governance active: permit validation, opt-out filtering, DP ($\epsilon=10$), audit logging.

The 6.6pp centralized-federated gap with Ditto demonstrates that privacy-preserving FL achieves clinically meaningful performance while maintaining full EHDS compliance. The governance layer’s negligible overhead validates the framework’s design principle: compliance should be a built-in capability, not a performance-degrading afterthought.

C. Comparison with Existing Approaches

Existing FL frameworks (Flower [22], NVIDIA FLARE [23]) provide robust FL infrastructure but lack governance integration. FL-EHDS is the only framework implementing: HDAB permit lifecycle, Article 71 opt-out enforcement, multi-HDAB cross-border coordination, and GDPR Article 30 audit persistence. The governance layer’s modular design allows integration as a wrapper around existing FL frameworks—e.g., as a Flower strategy plugin—enabling EHDS compliance within established ecosystems.

VI. IMPLEMENTATION ROADMAP

Table VI presents a phased roadmap aligned with EHDS milestones.

A. Stakeholder-Specific Recommendations

EU Policymakers: The March 2027 delegated acts represent the critical window. We recommend: (1) explicit guidance on gradient/model data status under GDPR; (2) standardized HDAB evaluation criteria for FL-based analytics; (3) technical specifications for FL within SPEs; (4) interoperability standards for cross-border HDAB communication.

National Authorities: The 2–3 year Nordic advantage [5] demonstrates that early investment in HDAB organizational capacity is essential. Priorities include: staff training on PET evaluation, coordination protocols with other Member States, and proactive citizen engagement about secondary use and opt-out rights. The current timeline risks a “two-speed” EHDS

TABLE VI
EHDS IMPLEMENTATION ROADMAP

Phase	Timeline	Priority Actions
Foundation	2025–26	Reference implementations; multi-MS pilots; FHIR acceleration
Clarification	2027	Delegated acts; PET legal guidance; HDAB standards
Scaling	2028–29	Production deployment; capacity building; citizen engagement
Operation	2029–31	Full cross-border analytics; genetic/imaging extensions

where citizens in well-prepared countries benefit from secondary use while others are excluded.

Healthcare Organizations: Preparation cannot wait for 2029. Organizations should: (1) accelerate FHIR compliance beyond the current 34% baseline [8]; (2) participate in Health-Data@EU pilots to gain operational experience; (3) assess computational infrastructure for FL participation; (4) develop internal governance policies for HDAB data access requests; (5) establish citizen communication strategies addressing opt-out provisions.

VII. DISCUSSION AND CONCLUSIONS

A. Key Finding: Governance Before Technology

Our systematic synthesis reveals that **legal and organizational uncertainties—not technical barriers—constitute the critical blockers** for EHDS implementation. While technical challenges (hardware heterogeneity, non-IID data, communication costs) are significant, they are tractable through known algorithmic solutions. In contrast, unresolved regulatory questions create compliance uncertainty that cannot be resolved through engineering. Without clarification of PET data status, organizations face potential GDPR violations regardless of technical privacy measures implemented.

This finding aligns with van Drumpt et al.’s [7] conclusion that governance frameworks are prerequisites, not alternatives, to technical solutions. The implication for EHDS implementation is clear: investment in governance capacity (HDAB staffing, cross-border protocols, citizen engagement) must precede or at least accompany technical infrastructure deployment.

B. Digital Health Equity Implications

The significant variation in HDAB capacity across Member States raises fundamental equity concerns. Forster et al. [9] document data access timelines ranging from 3 weeks (Finland) to over 12 months (France), reflecting deeply rooted differences in institutional capacity, digital health maturity, and governance culture. If Nordic countries operationalize secondary use by 2029 while Southern and Eastern European

states struggle with basic HDAB establishment, the EHDS risks creating a “data dividend” that benefits well-resourced research ecosystems while excluding others.

This two-speed implementation has several concrete consequences: (1) research datasets will be biased toward populations in data-ready countries, potentially missing genetic variants, disease patterns, and treatment responses characteristic of underrepresented populations; (2) healthcare organizations in less-prepared Member States will be unable to participate in collaborative FL studies, missing opportunities for model improvement on their patient populations; (3) citizens in well-prepared countries benefit from research insights derived from FL analytics, while others are excluded from the same benefits.

Our framework addresses this through: (a) the HDAB simulation backend enabling capacity building through training and pilots before production services are established; (b) graduated complexity levels allowing Member States to begin with simple permit workflows and progressively add cross-border capabilities; (c) comprehensive documentation and terminal-based interfaces requiring minimal infrastructure investment for initial deployment.

C. Citizen Trust and Transparency

The success of the EHDS ultimately depends on citizen trust. Article 71 opt-out mechanisms provide a formal right to refuse secondary use, but meaningful exercise of this right requires awareness and understanding. Van Drumpt et al. [7] demonstrate that public trust depends primarily on institutional transparency rather than technical privacy guarantees alone. Our framework supports transparency through: (1) audit trails accessible for citizen review upon request; (2) clear documentation of how data is processed within FL rounds; (3) per-purpose opt-out granularity allowing citizens to permit research use while blocking commercial applications.

The interaction between opt-out rates and FL model quality creates a tension: high opt-out rates reduce training data, potentially degrading model performance and widening health outcome disparities. Proactive citizen engagement—explaining how FL differs from data centralization, demonstrating governance safeguards, and providing transparent audit access—is essential for maintaining low opt-out rates while respecting individual autonomy.

D. Limitations

The governance framework is validated through simulation rather than binding to actual HDAB services, which do not yet exist. The evidence synthesis captures publications through January 2026; the rapidly evolving regulatory landscape may invalidate some findings. Our experimental validation uses public clinical datasets rather than production European EHR systems.

E. Conclusions

This paper presents an operational governance framework for the European Health Data Space, demonstrating that EHDS regulatory compliance is achievable through systematic mapping of Articles to implementation patterns. The framework’s

three-layer architecture integrates governance mechanisms, privacy-preserving analytics via Federated Learning, and data holder components with FHIR/OMOP interoperability.

Our key finding—that legal uncertainties and organizational capacity, not technical barriers, are the critical blockers—has direct policy implications. The March 2027 delegated acts represent a critical window for resolution. Without explicit guidance on PET data status, controller allocation, and model anonymity thresholds, the 2029 secondary use deadline arrives with EHDS adoption inhibited by governance uncertainty rather than technical limitations.

Future work: (1) integration with HealthData@EU Pilot infrastructure; (2) citizen attitude studies across diverse European populations examining opt-out intentions and trust factors; (3) economic sustainability modeling for HDAB operations; (4) longitudinal tracking of implementation trajectories to identify effective governance patterns.

ACKNOWLEDGMENTS

The author thanks Prof. Sadi Alawadi for supervision and guidance, and 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 EU*, L 2025/327, Mar. 2025.
- [2] A. Ganna, E. Ingelsson, and D. Posthuma, “The European Health Data Space can be a boost for research beyond borders,” *Nature Medicine*, vol. 30, pp. 3053–3056, 2024.
- [3] 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.
- [4] P. Quinn, E. Ellyne, and C. Yao, “Will the GDPR restrain health data access bodies under the EHDS?” *Computer Law & Security Review*, vol. 54, art. 105993, 2024.
- [5] 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.
- [6] H. Fröhlich et al., “Reality check: The aspirations of the EHDS amidst challenges in decentralized data analysis,” *J. Med. Internet Res.*, vol. 27, art. e76491, 2025.
- [7] 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, art. 1602101, 2025.
- [8] R. Hussein et al., “Interoperability framework of the EHDS for secondary use: Interactive EIF-based standards compliance toolkit,” *J. Med. Internet Res.*, vol. 27, art. e69813, 2025.
- [9] R. 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.
- [10] L. Svingel et al., “Shaping the future EHDS: Recommendations for implementation of Health Data Access Bodies,” *Eur. J. Public Health*, vol. 35, Suppl. 3, pp. iii32–iii38, 2025.
- [11] C. Christiansen et al., “Piloting an infrastructure for secondary use of health data: Learnings from the HealthData@EU Pilot,” *Eur. J. Public Health*, vol. 35, Suppl. 3, pp. iii3–iii4, 2025.
- [12] M. Shabani and P. Borry, “The European Health Data Space: Challenges and opportunities for health data governance,” *Eur. J. Human Genetics*, vol. 32, no. 8, pp. 891–897, 2024.
- [13] OHDSI, “The Book of OHDSI: Observational Health Data Sciences and Informatics,” 2019.
- [14] B. McMahan et al., “Communication-efficient learning of deep networks from decentralized data,” in *Proc. AISTATS*, pp. 1273–1282, 2017.
- [15] P. Kairouz et al., “Advances and open problems in federated learning,” *Found. Trends Mach. Learn.*, vol. 14, no. 1–2, pp. 1–210, 2021.

- [16] T. Li *et al.*, “Federated optimization in heterogeneous networks,” in *Proc. MLSys*, vol. 2, pp. 429–450, 2020.
- [17] L. Zhu, Z. Liu, and S. Han, “Deep leakage from gradients,” in *Proc. NeurIPS*, vol. 32, pp. 14774–14784, 2019.
- [18] I. Mironov, “Rényi differential privacy,” in *Proc. IEEE CSF*, pp. 263–275, 2017.
- [19] Z. L. Teo *et al.*, “Federated machine learning in healthcare: A systematic review,” *Cell Reports Medicine*, vol. 5, no. 2, art. 101419, 2024.
- [20] Z. Qu *et al.*, “FedLESAM: Federated learning with locally estimated sharpness-aware minimization,” in *Proc. ICML*, PMLR 235, 2024.
- [21] Y. Chen, X. Cao, and L. Sun, “HPFL: Hot-pluggable federated learning with shared backbone and personalized classifiers,” in *Proc. ICLR*, 2025.
- [22] D. J. Beutel *et al.*, “Flower: A friendly federated learning research framework,” *arXiv:2007.14390*, 2023.
- [23] NVIDIA, “NVIDIA FLARE: An open-source federated learning platform,” *GitHub Repository*, 2023.

APPENDIX

This appendix provides formal algorithmic descriptions demonstrating how EHDS regulatory requirements are operationalized at the implementation level.

A. HDAB Permit Validation (Art. 53)

Algorithm A.1: Data Permit Validation

Input: Permit P , round t , requested categories \mathcal{C}
Output: Boolean validity

```

// Check temporal validity (permit expiration)
if CurrentTime() > P.valid_until then
    raise PermitExpiredError
// Check purpose alignment (Article 53)
if P.purpose ∉ AllowedPurposes then
    raise PurposeMismatchError
// Check data category authorization
for each category  $c \in \mathcal{C}$  do
    if  $c \notin P.authorized\_categories$  then
        raise UnauthorizedCategoryError
// Log access for GDPR Article 30 compliance
AuditTrail.log(permit=P, round=t, categories= $\mathcal{C}$ )
return True

```

Validation checks include: temporal validity (explicit start/end dates), purpose alignment against Article 53 permitted categories, data category authorization (demographics, diagnoses, medications, genetic data require separate authorization), and comprehensive audit logging satisfying GDPR Article 30 for regulatory inspection.

B. Article 71 Opt-Out Registry Protocol

The protocol supports: (1) blanket opt-out (all secondary use); (2) purpose-specific opt-out (e.g., commercial applications only); (3) category-specific opt-out (e.g., genomics only). Registry caching with configurable TTL balances compliance (<10ms latency) with timely propagation of new opt-out decisions.

C. Cross-Border HDAB Consensus

Cross-border coordination implements: permit aggregation from each participating HDAB, real-time status synchronization, data residency compliance verification, and graceful degradation when one HDAB revokes a permit mid-study.

Algorithm A.2: Opt-Out Filtering

Input: Local dataset \mathcal{D}_h , purpose p , categories \mathcal{C}
Output: Filtered dataset \mathcal{D}'_h

```

// Synchronize with national opt-out registry (LRU-cached)
OptOutRecords ← FetchOptOutRegistry(MemberState)
 $\mathcal{D}'_h \leftarrow \emptyset$ 
for each record  $r \in \mathcal{D}_h$  do
    opted_out ← False
    // Check purpose-specific opt-out
    if (r.pseudonymized_id, p) ∈ OptOutRecords then
        opted_out ← True
    // Check category-specific opt-out
    for each  $c \in \mathcal{C}$  do
        if (r.pseudonymized_id, c) ∈ OptOutRecords then
            opted_out ← True
    if not opted_out then
         $\mathcal{D}'_h \leftarrow \mathcal{D}'_h \cup \{r\}$ 
AuditLog.record(total=| $\mathcal{D}_h$ |, filtered=| $\mathcal{D}'_h$ |)
return  $\mathcal{D}'_h$ 

```

Algorithm A.3: Multi-HDAB Coordination

Input: Study S , participating Member States \mathcal{M}
Output: Coordination status

```

permits ← {}
for each MS  $m \in \mathcal{M}$  in parallel do
     $P_m \leftarrow \text{SubmitPermitRequest}(\text{HDAB}_m, S)$ 
    permits[m] ← AwaitApproval( $P_m$ )
// Consensus: ALL HDABs must approve
if  $\exists m : \text{permits}[m] = \text{DENIED}$  then
    NotifyAll( $\mathcal{M}$ , “Study denied by ” + m)
    return DENIED
// Harmonize constraints across permits
 $P_{unified} \leftarrow \text{HarmonizeConstraints}(\text{permits})$ 
    // min(durations), min( $\epsilon$ -budgets), union(categories)
return APPROVED,  $P_{unified}$ 

```

Algorithm B.1: FHIR R4 Data Harmonization

Input: Raw EHR records \mathcal{R} , feature spec \mathcal{F}
Output: Training tensors (X, y)

```

// Detect source format (HL7 v2, CDA, CSV, etc.)
format ← DetectFormat( $\mathcal{R}$ )
parser ← GetParser(format)
records ← parser.parse( $\mathcal{R}$ )
// Map to standard terminologies
for each  $r \in \text{records}$  do
    r.diagnoses ← MapToICD10(r.diagnoses)
    r.medications ← MapToATC(r.medications)
    r.labs ← MapToLOINC(r.labs)
// Convert to FHIR R4 and validate
fhir_bundle ← ToFHIR(records)
ValidateFHIR(fhir_bundle)
// Extract ML-ready tensors
 $X \leftarrow \text{ExtractFeatures}(\text{fhir\_bundle}, \mathcal{F})$ 
 $X \leftarrow \text{StandardScaler.fit\_transform}(X)$ 
 $y \leftarrow \text{ExtractLabels}(\text{fhir\_bundle})$ 
return  $(X, y)$ 

```

The pipeline supports 6 FHIR R4 resource types (Patient, Observation, Condition, MedicationRequest, Procedure, DiagnosticReport) with standard coding systems (SNOMED-CT, LOINC, ICD-10, ATC, UCUM). Given that only 34% of European providers achieve full FHIR compliance [8], this preprocessing is essential for practical EHDS deployment.

OMOP CDM Integration: An alternative harmonization path via OMOP CDM v5.4 supports observational research networks (EHDEN, OHDSI). ETL pipelines transform source EHR to OMOP, enabling consistent feature engineering via the FeatureExtraction package.

Beyond FHIR R4, the framework implements IHE (Integrating the Healthcare Enterprise) profiles for secure, auditable cross-border exchange:

ATNA (Audit Trail and Node Authentication): TLS mutual authentication between FL nodes; syslog audit messages (RFC 5424) for all data access events. Maps directly to GDPR Article 30 record-keeping.

BPPC (Basic Patient Privacy Consents): Maps Article 71 opt-out decisions to BPPC consent documents; XDS.b integration for consent sharing across institutions; consent enforcement at FL training initiation.

XCA (Cross-Community Access): Cross-border document query/retrieve via Initiating/Responding Gateways; patient identity correlation across communities using pseudonymized identifiers.

XUA (Cross-Enterprise User Assertion): SAML 2.0 assertions for federated authentication; role-based access control; HDAB authorization token propagation across Member States.

Security: eIDAS-compliant electronic signatures for permits, TLS 1.3 for all cross-border communication, certificate-based node authentication within the EU trust framework. Metadata follows DCAT-AP Health extension for dataset cataloging with W3C PROV-O provenance tracking.

D. Differential Privacy Mechanism

Algorithm D.1: Gaussian DP with Rényi Accounting

Input: Gradient Δ , clip norm C , budget ϵ , δ

Output: Noisy gradient $\tilde{\Delta}$

$\sigma \leftarrow C \cdot \sqrt{2 \ln(1.25/\delta)}/\epsilon$

for each parameter $w \in \Delta$ **do**

$\tilde{w} \leftarrow w + \mathcal{N}(0, \sigma^2)$

PrivacyAccountant.spend(ϵ) // Rényi DP composition

if budget_exhausted() **then abort** // Hard privacy stop

return $\tilde{\Delta}$

Rényi DP (RDP) provides $5-6\times$ tighter composition bounds than naive composition for the 20+ rounds typical of EHDS cross-border studies, enabling longer training within equivalent privacy guarantees.

E. Secure Aggregation

Pairwise masking with ECDH key exchange ensures the aggregation server observes only the sum of gradients, never individual hospital contributions. Shamir's (t, K) -threshold secret sharing provides dropout resilience: if fewer than t

clients complete a round, reconstruction fails gracefully without privacy leakage.

The open-source codebase ($\sim 40,000$ lines, 159 Python modules) provides:

- **Core** (36+ modules): 17 FL algorithms, secure aggregation, differential privacy, 6 Byzantine resilience methods, HDAB governance APIs
- **Data** (7 modules): FHIR R4 loader, OMOP harmonizer, 5 clinical dataset loaders
- **Dashboard** (15 modules): Streamlit web interface with EHDS governance workflow, real-time FL monitoring, permit management
- **Terminal** (15 modules): Professional TUI with 11 specialized screens for algorithm configuration, dataset management, privacy settings, cross-border coordination
- **Benchmarks:** Reproducible experiment suite generating all paper results (105 experiments with checkpoint/resume support)
- **Tests:** Unit and integration tests for governance, DP, configuration, orchestration

Repository:

[https://github.com/FabioLiberti/](https://github.com/FabioLiberti/FL-EHDS-FLICS2026)

FL-EHDS-FLICS2026