Evidence and Position Paper
AI-POWERED SCALING OF THE EHDS

From Compliance to Value:
Multi-Stakeholder Consultation and calls to action

Based on four stakeholder roundtables (patients, hospitals, decision-makers, industry)
i~HD and the AIDAVA Consortium, June 2026

1. The Policy Question

The EHDS Regulation (EU 2025/327) is a landmark achievement: it establishes the citizen rights, the governance, and technical baseline for a European health data space. As such, the Regulation provides the organisational scaffolding and the rules of access; the data will be supplied, and ultimately given value, by the community that produces and uses health data: clinical users, hospitals and registries, innovators, regulators, and the Health Data Access Bodies across 27 Member States.

Four stakeholder roundtables convened by the AIDAVA consortium in 2026 with patients, hospitals, decision-makers and industry, converged on the same finding: regulatory compliance and value extraction are two very different things. Member States will meet the 2029 deadlines. Whether the data exchanged will be clinically trustworthy, semantically consistent, of a fit-for-purpose quality, efficiently provided and economically usable for research and innovation is an entirely different question and one the current implementation agenda does not yet adequately answer. Two questions remain unanswered:

- How fit for purpose will the data be? and
- At what recurring cost to society as a whole?

This briefing argues that AI-assisted curation, demonstrated by the AIDAVA project, offers a credible answer to both — and that specific enabling conditions must be established for that answer to scale.

2. Eighteen Years of Investment: What We Have, What Is Still Missing

By 2027 the EU will have invested over **€370 million** in health data interoperability since 2008, approximately three quarters of it supporting EHDS1 primary use infrastructure and the remaining one quarter focused on secondary use. This investment has produced a genuine achievement: a coherent, regulated interoperability baseline in the EEHRxF, an agreed specification, a growing vendor ecosystem, and momentum across Member States. That is the foundation.

What remains unresolved, confirmed in several stakeholder consultations, are three structural challenges that go well beyond regulatory compliance:

Funded by the European Union

AIDAVA - Horizon Europe 10105706

- **Building trust in the data exchanged.** Data quality and semantic interoperability at the source remain poor. Approximately 80% of hospital health data exists as unstructured narrative free text, 30% of contain redundancies, 40% contain errors or inconsistencies, 10% contain potentially life threatening errors[1]. Exporting this data into the EHDS, without addressing quality at the source, exports the problem at European scale.
- **Complexity and scalability.** The policy to deploment timeline of the step-by-step priority data category approach for EHDS1 is 8–15 years. A genuinely comprehensive EHDS would be decades away on the current trajectory. There is no mechanism to accelerate this without a different approach to interoperability and data quality at the source.
- **Cost to data holders and to society**. We have no sustainable funding model for EHDS readiness of health data. Most hospitals are in survival mode, focused on financial consolidation, lacking both the budget and the bandwidth to invest in data infrastructure. Meanwhile, the downstream cost of poor data quality falls on every data user, every year, and ultimately on patients and health systems through costs of innovative technologies for health systems.

EHDS compliance is necessary but not sufficient to achieve the intended benefits. Success will be measured not in EHDS Regulation compliance per se, but in how this compliance impacts trustworthy, well-structured, re-usable, **value-creating data** in routine care and research. Clinical utility, semantic accuracy across countries, affordable cost, and market adoption are the dimensions on which the EHDS success will ultimately be judged.

3. The Data Curation Cost Burden: Who Pays, How it Impacts Patients

The EHDS Impact Assessment focused on the costs of infrastructure and governance. It did not address a different category of cost, one that has been traditionally large and expected to grow as EHDS2 scales: the cost borne by data users, pharmaceutical companies, biotech firms, academic researchers, health technology innovators to curate and prepare non-interoperable, suboptimal quality European health data from the data holders before any analysis can begin. This cost does not disappear with EHDS2; the flow is easier, the quality of the data remains the same and the processing cost is ultimately passed through to medicine prices, health technology costs and ultimately to the public purse.

European pharma currently spends approximately **€1.2 – 1.4 billion** per year preparing health data for research before any analysis begins, a figure projected to exceed **€2.5 billion by 2030** (MarketsandMarkets, 2024). Industry surveys consistently find that 50–80% of this expenditure goes not on generating insights but on cleaning, harmonising, and structuring data. The industry roundtable confirmed that this preparation burden is the most critical bottleneck to pharma investment in EU clinical research. Participants stressed

[1] https://mydata.org/2024/05/27/as-a-citizen-or-patient-the-main-reason-to-be-in-charge-of-my-health-data-is-to-ensure-its-quality-and-continuity-of-care/

that EHDS will only attract industry investment if it demonstrably enables faster, cheaper, more targeted clinical trials in the EU — a market currently at a plateau. The data curation industry will not disappear; it will be transformed. When data quality is addressed at source, the industry's work shifts from low-value repeated data preparation to high-value analysis and knowledge generation across data spaces. That is an outcome the whole ecosystem should welcome.

The data curation cost burden

The data curation cost burden has a direct consequence for Europe's competitiveness. Europe's share of commercial clinical trials dropped from 22% in 2013 to 12% in 2023, even as the total number of trials increased by 38% globally (IQVIA, October 2024). EFPIA estimates this translates to 60,000 fewer patients accessing a trial involving a country within the EEA,, with European clinical trials hampered by a slow and fragmented research ecosystem.

4. The AI-Powered Alternative: Curate Once, Use Many Times

The AIDAVA project (1) has developed and validated at TRL 5–6 a prototype that directly addresses both the fitness gap and the cost burden. AIDAVA does not claim to be the only or final solution; it has advanced and tested the concept and technologies to a maturity level that allows deriving concrete recommendations on the way forward.

How it works: Heterogeneous records from multiple hospital systems - each with different formats, coding systems and quality levels - are ingested and curated once into a semantically structured Personal Health Knowledge Graph (PHKG) aligned with a Simplified Upper Level Ontology (SULO). AI tools automatically curate the data, correct inconsistencies and resolve gaps; where automation is insufficient, a conversational interface engages the patient or a professional curator, based on the level of digital health literacy required to answer the question. Once built and maintained with updated information from the patient record, the PHKG serves as the single quality-assured source from which EEHRxF exports and research datasets are generated on demand via SPARQL queries. The principle is simple: **curate once, use many times.**

What this means for EHDS1, EHDS2 and clinical trials

For EHDS1: A FHIR Implementation Guide is transformed into a SPARQL query which is executed on the PHKG to generate the corresponding EEHRxF-compliant exports on demand. This complements, without replacing, existing MyHealth@EU infrastructure, adding clinical depth and semantic accuracy to what is currently exchanged.

For EHDS2: HDAB-facilitated research queries are transformed through an AI supported assistant into a SPARQL query executed on the PHKG to extract the required data, at a fraction of the manual extraction cost. This data is then mapped into the required target format to produce structurally consistent data rather than raw records. The PHKG , integrating ALL data from each individual patient, also resolves the data enrichment challenge identified by TEHDAS2: multi-source record linkage that would normally require complex per-study governance and processing is done once at the point of curation, and inherited by all downstream uses.

For clinical trials: the EHR-to-EDC pipeline, identified by the industry roundtable as the highest-value entry point for pharma, becomes automated. Site eligibility, patient matching, and trial data collection RWD Generation can draw directly from the PHKG, reducing per-trial data preparation costs and accelerating recruitment across EU sites. This is the lever that could reverse the plateau in European clinical trial volume.

For patient and care providers: Better data for better care - less burnout.

Evidence from deployment

AIDAVA has been validated across four hospital settings in three countries (Netherlands, Estonia, Austria) and three languages, around two clinical use cases: a cardiovascular risk scoring system generating SMART 10-year CVD risk scores and a federated breast cancer clinical registry supporting physicians to check metrics/queries across site. In addition, AIDAVA automatically generates the IPS in HL7 FHIR. The peer-reviewed JMIR Medical Informatics paper (2) demonstrates the framework's ability to dynamically detect completeness and consistency issues throughout the data lifecycle using SHACL validation, the technical foundation for the data quality at source claim. External validation through the eCREAM project provides independent confirmation of AIDAVA's data quality assessment and correction capabilities beyond the consortium's own evaluation.

The cost comparison that matters

The AIDAVA draft cost model estimates the total EU-27 cost of implementing AI-assisted curation at scale, covering not only curation for all EU records, but also EHDS1 EEHRxF generation of ± 1,9 EEHRxF/citizen/year and 540 K queries/year EHDS2 query responses, at approximately **€1.04 billion over five years.** The cost for EHDS1 and EHDS2 only for the same volume was estimated under the AIDAVA draft cost model at **€6.6 billion over five years)**

The comparison in plain terms

The total cost of deploying AIDAVA based AI curation for all EU patients across EU-27 over five years is **less than one year of what European innovators currently spend just on preparing health data** before research can begin. That is the investment required to eliminate the structural inefficiency — permanently. The cost is borne once; the savings accrue to innovators, health systems, and ultimately patients, every year thereafter.

Note: the AIDAVA cost model parameters are subject to national-level validation. The comparison is indicative; the direction of travel is robust.

5. What Needs to Change: Governance, Infrastructure, Business Model

All four roundtables converged on a finding that goes beyond the technology demonstration: the conditions for scaling AI-assisted health data curation do not yet exist. Three structural gaps must be addressed simultaneously

A European Semantic Foundation

Any AI-assisted interoperability strategy at EU scale depends on one enabling condition currently absent from the EHDS implementation roadmap: a shared semantic reference layer to which hospitals map their data fields once and to which clinical standards map their concepts[2]. Without this layer, every hospital, every project, and every Member State independently solves the same terminology mapping problem, relating local clinical concepts to SNOMED, LOINC, ICD, HL7 FHIR, OMOP and other standards. The result is expensive, duplicated effort producing inconsistent results across borders.

The AIDAVA project has developed **SULO**, the Simplified Upper Level Ontology, as an initial answer. SULO is a minimal set of classes and relations guiding the development of Personal Health Knowledge Graphs. It is published under a CC0 public domain licence, a genuine public good, not a proprietary standard. It is minimal by design: **a lightweight bridge layer, not a replacement,** for SNOMED, LOINC, ICD or HL7 FHIR, but a common set of well declared semantic principles that enable semantic alignment[3] and making cross-standard alignment automatic once mappings between the standards and SULO exist. SULO has been tested in AIDAVA and the Res-Q+ stroke registry project and its peer-reviewed specification was published at JOWO/FOIS 2025 (3); additional work is needed for mapping across existing standard.

[2] A similar approach is already implemented across Swiss hospitals through the SPHN initiative; a major burden for the data holders remain the cost of curation from their current raw data into the SPHN semantic representation

[3] Existing de facto standards like FHIR, openEHR, OMOP are not always semantically aligned. An event in one standards can be a finding in another one, requiring interpretation, and therefore potential loss in interoperability, by each "human mapper"

The policy ask is a commitment to the principle that the EU will govern and maintain a shared simplified upper-level ontology for health as a public asset, and that Standards Development Organisations - HL7, SNOMED International, CEN/TC 251, openEHR, OMOP - will be formally engaged to develop mappings to it. The industry roundtable confirmed that without this governance, the AIDAVA PHKG risks becoming yet another isolated “common” data representation. A formal governance structure, EU-supported and maintained by one or more SDOs, is the condition on which the entire scalability argument depends.

5.2 A Sustainable Infrastructure Funding Model

Project-based funding is insufficient. The investments in improving the health data infrastructure should not be made piecemeal, initiative by initiative nor can they be funded by single beneficiaries such as Pharma. The RT partifcipants called for a shift toward treating health data infrastructure as a pre-competitive national asset co-funded by Member States, industry, and research bodies. A clear economic model with ROI articulated for each stakeholder group (hospitals, pharma, regulators, patients) must be developed and presented to decision-makers and relevant EC DGs.

Three distinct funding challenges must be addressed:

- Data holders (hospitals and other health care providers) have no reimbursement model for generating EHDS ready data at the source. The cost of curating the data, generating EEHRxF exports, responding to HDAB queries, and maintaining data quality could not fall on clinical institutions already under severe financial pressure.
- Public investment must be the primary driver. National funding, not industry alone, must underpin the infrastructure. Industry will contribute when the business case is demonstrated; it will not invest as the primary funder of public infrastructure. A well-designed incremental adoption roadmap, starting with digitally mature large hospitals and scaling to smaller institutions and general practice, is more likely to deliver early wins that inspire scale-up.
- The business case must be made for each stakeholder group.
    - For pharma, the EHR-to-EDC clinical trial use case is the entry point.
    - For hospitals, reduced data extraction cost and direct benefit from data such as less burnout of physicians, reduces time in retrieving patient information, improved patient outcomes, AI readiness.
    - For health systems and patients, lower medicine prices and faster access to innovation.

These cases must be developed with measurable ROIs and presented to the relevant EU and national ministries.

5.3 Workforce training to build new type capacity

The roundtables noted a significant and widespread misconception: that AI will solve the interoperability problem without requiring clear semantic foundations. Building awareness of how these technical pieces fit together and developing the workforce to deploy and govern them, is as important as the technology itself.

6. Calls to Action

These calls to action reflect the consensus of four stakeholder roundtables. They address four interdependent conditions for success: technology development with demonstrated **feasibility**, governance that ensures EU wide adoption and **scalability**; financial evidence of **sustainability** and long term ; capacity that secures **accessibility** ; and **coordination** that makes it coherent with the broader EHDS implementation agenda..

1. Technology: Test and expand

Support additional clinical use cases and national implementation pilots, involving multiple data holders, real governance frameworks, real data access use cases, in at least 3–5 Member States through dedicated EU funding calls. Recommended priority entry points are electronic health record (HER)-to-electronic data capture (EDC) clinical trial enablement and patient matching for pharma; federated registry population for clinical research.

2. Governance: A European health data semantic strategy

Establish an EU Health Data Semantic Strategy including: (a) governance and sustainable public funding for a shared Simplified Upper Level Ontology (SULO or equivalent) as an EU public asset; (b) formal engagement of HL7, SNOMED International, CEN/TC 251 and other int'l Standards Development Organisations to develop mappings based on agreed, shared semantics and to enable cost-effective, AI-powered mapping; and (c) open licensing for all EU-funded semantic assets

3. Preparation: Capacity building and workforce preparation

Invest in tools and the needed set of skills to enable this change; etsablsih an EU open-source ecosystem, ensuring shared tools and semantic assets are accessible to all Member States regardless of their starting point.

4. Infrastructure: Define a Sustainable Data Infrastructure Funding Model

A paradigm shift is needed toward treating health data infrastructure as a pre-competitive national asset — co-funded by public Member States/EU budgets, industry and research bodies. A clear economic model with ROI articulated for each stakeholder group (hospitals, pharma, regulators, patients) must be developed and adopted by decision-makers.

5. Coordination: EU and national action aligned with the regulatory framework

Coordinated EU and national action is needed, in line with the AI Act, MDR, and the Biotech Act's data quality accelerator provisions, with strong involvement of Standards Development Organisations as governance partners. The EHDS Board, eHealth Network, and AI Office should establish a joint regulatory learning mechanism, including AI Act regulatory sandboxes, to resolve open qualification and governance questions for AI-enabled health data infrastructure.

Publications

(1) AIDAVA — Horizon Europe GA 101057062, €8.6M, 14 partners, 9 countries, coordinated by Maastricht University, September 2022 – August 2026.

(2) Declerck et al., JMIR Medical Informatics, November 2025.

(3)SULO specification: JOWO/FOIS 2025, Catania.

Prepared by the AIDAVA Consortium & i~HD (Institute of Innovation through Health Data) | June 2026 | www.aidava.eu