Reframing the European Health Data Space: From Digital Plumbing to a 12-digit value Strategic AI Asset

A. Aldo Faisal*

Imperial College London, UK

Universität Bayreuth, Germany

April 16, 2026

Executive Summary: The European Health Data Space (EHDS) is currently valued primarily as an administrative infrastructure project, with projected benefits of roughly €11 billion over a decade. This paper argues that this assessment drastically undervalues the initiative by ignoring the emergence of Foundation Models (FMs). By reframing the EHDS as the data substrate for sovereign AI, we demonstrate through rigorous estimation that the true strategic value is likely an order of magnitude higher—approaching **€90 billion annually** at maturity, with a ten-year Net Present Value (NPV) of approximately €350–400 billion. This valuation accounts for healthcare efficiency gains, and intersectoral healthcare shift lead to preventive medicine and the creation of health foundation models such as Nightingale AI enabling medical solutions not imaginable prior to AI, which become viable at the scale of half a billion lives.

1 Introduction

The European Union's flagship plan for health data, the European Health Data Space, is usually described in the sober language of infrastructure. It promises interoperable records, smoother cross-border care and a tidier framework for researchers seeking access to medical data. In the Commission's own impact assessment, the numbers attached to this are modest: roughly €11bn of benefits over a decade, largely from better plumbing—cleaner pipes for the flow of information.

That view is already out of date. Once you place the EHDS in the world of ChatGPT-like foundation models, the scale of what is at stake changes by at least an order of magnitude. Instead of being a digital filing system with some research benefits on the side, it becomes the fuel source for sovereign, health-specific AI models trained on the lived experience of more than 450m people. And once you model that synergy, even with conservative assumptions, you end up in a very different place.

*aldo.faisal@imperial.ac.uk

The Baseline: The "Plumbing" Business Case The conventional business case for the European Health Data Space (EHDS) relies on two pillars: Primary Use, which concerns direct healthcare delivery, and Secondary Use, covering research, innovation, and policy. In the sober language of infrastructure, the European Commission's own impact assessment attributes modest savings to this initiative: approximately **€11 billion over ten years (2025–2035)**.

These savings are largely derived from improved "plumbing"—cleaner pipes for the flow of information. The Commission identifies three primary sources of value. First, efficiency gains are expected from better access to health data, which reduces the duplication of medical tests—estimated at roughly 10% of imaging and laboratory tests—and enables faster prescription refills. Second, the initiative projects a 20–30% expansion in the digital health sector, driven by the creation of a unified market for Electronic Health Record (EHR) systems and wellness applications. Third, the Life Sciences and MedTech industries are expected to benefit from lower costs and reduced timeframes for acquiring real-world evidence (RWE) via *HealthData@EU*.

While these figures are supported by the TEHDAS Joint Action and demonstrate the technical feasibility of cross-border data sharing, they provide only a limited view. They value the infrastructure but fail to account for the strategic asset flowing through it. To derive a strategic valuation, we must incorporate the pace of innovation in artificial intelligence.

The Missing Variable so far: The AI Multiplier Once the EHDS is situated within the context of AI foundation models, the scale of the economic opportunity changes by at least an order of magnitude. Instead of serving merely as a digital filing system, the EHDS becomes the fuel source for sovereign, health-specific AI models trained on the lived experiences of more than 450 million people.

A back-of-the-envelope calculation illustrates the discrepancy. European healthcare spending was approximately **€1.72 trillion in 2023** (Eurostat), accounting for roughly 10.4% of EU GDP. Serious estimates of the impact of artificial intelligence on this expenditure are substantial. A study by MedTech Europe and Deloitte suggests that AI could ultimately save between €170 billion and €210 billion annually in European health systems—approximately 12% of total expenditure—through fewer unnecessary tests, shorter hospital stays, and more efficient use of scarce staff time. In parallel, the McKinsey Global Institute estimates the global value of generative AI to the pharmaceutical and medical products industry at between $60 billion and $110 billion annually.

Crucially, none of these high-value estimates rely on the EHDS; they describe what AI could achieve in a world of patchwork data and proprietary silos. The strategic question is how these figures evolve when a genuinely European, harmonised data layer is introduced.

The Core Argument: Scale as a Moat The technical reality of Foundation Models in healthcare, such as Nightingale AI or Delphi-2M, is that they do not improve linearly with data volume. Instead, they exhibit emergent capabilities that are unlocked only at massive scales. A model trained on 6 million lives—the scale of Singaporean data—is not merely "smaller" than one trained on 450 million lives; it is functionally inferior.

This distinction underpins the "Data Hegemony" thesis. The future of healthcare is intersectoral, achieved through digitisation that builds a shared service fabric across providers, social care, and payers. In this race, most competitors are trapped in a "Small Data" reality. US providers such as Kaiser Permanente possess high-quality but siloed data (∼12.5 million members) that is biased toward insured demographics. Conversely, digital leaders like Singapore or Denmark offer perfect depth but insufficient volume (∼6 million people). These models hit a "learning ceiling" where they lack sufficient examples to identify rare diseases or complex comorbidities.

The EU advantage is defined by volume and diversity. The EHDS offers access to approximately 450 million longitudinal records, which is two orders of magnitude larger than most national datasets. Furthermore, genetic diversity reduces algorithmic bias and significantly improves the model's ability to generalise globally.

The Science of Scale: What Evidence We Can (and Cannot) Borrow from Language Models for Health Models The central claim is not that "more data" automatically yields "better AI," but that *effective scale* changes what is feasible. In healthcare, high-value tasks are dominated by the long tail (rare diseases, rare subtypes, rare drug–gene–comorbidity interactions) and by the need for robust subgroup calibration across heterogeneous care pathways. The EHDS is therefore best interpreted as a step-change in *usable* evidence: a continent-scale substrate that, *conditional on implementation quality*, can materially expand tail coverage and cross-system generalisation beyond what most national or single-provider datasets can support.

For clarity, we distinguish the nominal population size from an *effective* training and validation scale. Let $N_{\text{eff}} = N \times f_{\text{usable}}$, where f_{usable} captures the fraction of records that are longitudinally linkable, semantically interoperable, sufficiently complete, and operationally accessible under EHDS governance (secure processing environments, standardised metadata, and consistent quality controls) [3, 12, 13]. Table **??** provides a heuristic "scaling ladder": it is not a claim of universal thresholds, but an intuition pump for why continental-scale *usable* data can enable qualitatively different reliability regimes for selected clinical tasks.

The central claim is not that "more data" automatically yields "better AI," but that *effective scale* changes what is feasible. In healthcare, high-value tasks are dominated by the long tail (rare diseases, rare subtypes, rare drug–gene–comorbidity interactions) and by the need for robust subgroup calibration across heterogeneous care pathways. The EHDS is therefore best interpreted as a step-change in *usable* evidence: a continent-scale substrate that, *conditional on implementation quality*, can materially expand tail coverage and cross-system generalisation beyond what most national or single-provider datasets can support.

For clarity, we distinguish the nominal population size from an *effective* training and validation scale. Let $N_{\text{eff}} = N \times f_{\text{usable}}$, where f_{usable} captures the fraction of records that are longitudinally linkable, semantically interoperable, sufficiently complete, and operationally accessible under EHDS governance (secure processing environments, standardised metadata, and consistent quality controls) [3, 12, 13]. Table **??** provides a heuristic "scaling ladder": it is not a claim of universal thresholds, but an intuition pump for why continental-scale *usable* data can enable qualitatively different reliability regimes for selected clinical tasks.

In the foundation-model literature, performance often follows predictable scaling relationships: loss and downstream capability improve approximately as power laws in model size, data, and compute [8, 9]. At the same time, many researchers report *emergent* abilities that appear discontinuously when models pass certain thresholds, although the phenomenon is debated and can be sensitive to measurement choices [10]. In image generation, diffusion-model quality similarly improves markedly with scale and training regimes, enabling high-fidelity synthesis at large dataset sizes [11].

Healthcare is not language, and we should be cautious when borrowing claims from LLM scaling. Nonetheless, two transferable lessons are robust enough to inform valuation.

First, marginal predictive gains can have non-marginal economic effects at the population scale. Even if average improvements are smooth, a small increase in calibration or discrimination in a high-cost pathway (e.g., avoidable admissions, complications, progression) can translate into large absolute savings when deployed across tens or hundreds of millions of patient-years.

Second, tail coverage drives practical utility. Many high-value medical applications are limited by the availability of sufficiently many diverse examples under consistent governance—not by model architecture alone. Larger, more heterogeneous datasets expand the set of reliably learnable subgroups, reduce brittleness under distribution shift, and improve the feasibility of counterfactual modeling for *selected pathways* (e.g., treatment-response heterogeneity in well-defined cohorts).

These observations motivate the two-layer valuation below: an efficiency layer that plausibly exists without EHDS-scale models, and a transformation layer whose feasibility and reliability are substantially amplified by EHDS-scale *usable* data.

These observations motivate the valuation approach below: an efficiency layer that plausibly exists without EHDS-scale models, and a transformation layer that plausibly requires (or is substantially amplified by) EHDS-scale data.

2 Methodology: A Two-Layer Value Stack

We adopt a social planner perspective over a ten-year horizon. We compare a **Baseline Scenario** (fragmented national data availability and heterogeneous access procedures) against an **EHDS+FM Scenario** (harmonised access procedures, interoperable data formats, and EU-grade foundation models). We deliberately separate two layers:

- **Layer 1 (Efficiency):** Automation and decision support that reduce waste in existing processes.
- **Layer 2 (Transformation):** New value streams that are either (i) unlocked by scale (e.g., prevention at acceptable false-positive rates), or (ii) enabled by EU-grade model generalisation and trust (e.g., export/licensing of clinical intelligence).

The distinction between Layer 1 and Layer 2 is not cosmetic; it corresponds to two qualitatively different sources of value.

Implementation and governance constraints mean that the data value is conditional. The EHDS does not imply unrestricted pooling of raw personal data. The Commission's communication emphasises that secondary-use processing is intended to occur in **secure processing environments** with strong privacy and cybersecurity guarantees, and that personal data should not be downloadable from such environments [3]. TEHDAS workstreams provide concrete operational guidance on data quality frameworks and architectures required for HealthData@EU to function at scale [12, 13]. Our valuation should therefore be read as *conditional on implementation quality* and *bounded by governance design choices.*

2.1 Efficiency Calculations (Layer 1)

Layer 1 (Efficiency) aggregates gains that largely improve existing workflows: reduced duplication, faster administrative throughput, and higher productivity in analytics and R&D. These mechanisms are plausible even in a world without an EHDS, but they are materially amplified by EHDS interoperability and governance-backed portability. In other words, Layer 1 is a conservative *floor*: it captures value from deploying AI at scale in health systems and life sciences, and then asks how much additional uplift is unlocked by an EHDS-grade data substrate. I.e. we are estimating the value of "doing things that we used to do by hand using AI".

We use a generic valuation form to structure assumptions:

$$V_{\text{eff}} = B_{\text{spend}} \times I_{\text{AI}} \times S_{\text{FM}} \times U_{\text{EHDS}}. \tag{1}$$

Here, B_{spend} is the relevant spend or market baseline, I_{AI} is the share plausibly addressable by AI, S_{FM} is the fraction of AI impact that depends on foundation-model capabilities (e.g. generative workflows, representation learning), and U_{EHDS} is the incremental uplift attributable to EHDS-scale access/interoperability (rather than AI per se). We report *gross* values and treat adoption and implementation costs separately in the sensitivity section.

Stream A: Healthcare Delivery (Efficiency & Quality) European current healthcare expenditure was approximately **€1.72 trillion in 2023** [4]. We take a conservative AI impact band of 5–12% (noting that some published figures include opportunity costs) and assume that 40–60% of that impact is mediated by FM-like capabilities. We then apply an EHDS uplift of 30–40%, representing the additional value of diversity, portability, and cross-border robustness enabled by EHDS governance and interoperability.

$$\text{Central Result:} \quad 1{,}720 \times 8.5\% \times 50\% \times 35\% \approx \textbf{€25.6 billion/year}.$$

Stream B: Data Reuse and Analytics Productivity The European Commission estimates that health data reuse is worth around **€25–30 billion annually** and could reach **€50 billion within 10 years** [3]. We interpret €50 billion/year as a plausible maturity-level baseline for the secondary-use economy under "Plain EHDS" (improved access procedures, larger addressable market, and reduced duplication in real-world evidence generation).

The incremental value considered here is a productivity uplift from foundation-model-enabled workflows *on top of* that baseline. Concretely, FMs can reduce the marginal cost and latency of (i) data discovery and feasibility queries, (ii) cohort definition and phenotype engineering, and (iii) repetitive analytic and reporting pipelines by turning previously bespoke integrations into reusable, governed utilities. The EHDS matters because interoperability, metadata standardisation, and common quality frameworks reduce the amount of rework required per dataset and per country, increasing the reusability of pipelines and the portability of learned representations [12, 13].

We therefore apply an FM-driven uplift band of 30–70% to the €50 billion/year maturity baseline. This does *not* assume that data cleaning disappears; it assumes that standardisation and common access patterns reduce the repeated fixed costs of "starting from scratch" for each new study, registry, or business analytics build.

$$\text{Central Result:} \quad 50 \times 50\% \approx \textbf{€25 billion/year}.$$

Stream C: Pharma and MedTech R&D Acceleration For life sciences, we take the McKinsey-cited MGI estimate of **$60–110 billion** annual generative-AI value as a global magnitude reference [7]. We assume an EU-relevant share and focus only on the fraction plausibly dependent on rich longitudinal real-world evidence (RWE) access. Under central assumptions, this yields:

$$\text{Central Result:} \quad \text{EU-relevant GenAI value } (\sim \text{€20–30 bn}) \times 30\text{–}50\% \approx \textbf{€10 billion/year (central)}.$$

.

$$\text{Central Result:} \quad 25 \times 40\% \approx \textbf{€10.0 billion/year}.$$

While Generative AI aids molecular discovery("finding the drug"), the economic bottlenecks in pharma lie downstream: in clinical Development (running trials which 95% 'discovered' drugs fail) and Market Access (patients getting reimbursed). The EHDS addresses both, effectively compressing the time-to-market and time-to-peak-sales.

For example, clinical Trial Optimization (Development), patient recruitment accounts for nearly 30% of clinical trial time; 80% of trials fail to meet enrollment deadlines. The EHDS-enabled foundational AI turns the EU into a single site for feasibility queries, allowing sponsors to identify rare patient phenotypes instantly across 27 nations. Furthermore, EHDS-scale historical data enables Synthetic Control Arms, reducing the need to recruit placebo patients and lowering trial costs by an estimated 15–20%.

The biggest financial risk for pharma (and medtech in a smaller scale but still applicable) today is reimbursement, i.e. proving to payers that the drug works better than the cheap generic. Rendering the EHDS useful for foundation models provides the digested "Real-World Evidence" (RWE) required for Value-Based Pricing. This is distinct from discovering the molecule. It is about the commercial survival of the innovation. Novel treatment launch success today depends on proving "Value" to payers (e.g., statutory insurance). This requires Real-World Evidence (RWE) showing that a therapy works in diverse, real-world populations, not just sterile trial settings. The EHDS provides the longitudinal substrate to generate this evidence without building bespoke registries for every country.

Therefore, factoring these two additional elements (that due to lack of a European-scale AI could not be factored in the McKinsey numbers) we believe that estimate is converative, as we move beyond the improved rate of discovery, the RWE and trial-optimization portion of the value chain that accelerating development timelines and securing faster reimbursement, we estimate a central impact of €10 billion annually (which is conservative)

2.2 Strategic Upside of AI valuation calculation (Layer 2)

Layer 2 reflects mechanisms that are constrained by scale, calibration, and generalisation. These are the channels where the EHDS is not just a multiplier but potentially a *prerequisite*. The two mechanisms listed ("Shift Left" and export/licensing) are deliberately framed as parameterised opportunity magnitudes: they are sensitive to service design, regulation, and industrial strategy, and therefore should not be interpreted as guaranteed savings. They represent the payoff to implementing EHDS well enough that continental-scale modelling becomes operationally usable, trusted, and reusable. This layer estimates doing things we could not do before AI.

The efficiency calculations above represent a conservative floor. A health foundation model trained and validated across EHDS-scale data may unlock two higher-order mechanisms that are limited by smaller datasets and fragmented governance.

Mechanism 1: "Shift Left" to Prevention and Monitoring The fundamental economic barrier to preventive medicine at the population scale is false positives: if prediction is noisy, screening large populations generates costly cascades. The strategic hypothesis is that EHDS-scale modelling improves calibration and subgroup performance sufficiently to make selected preventive pathways cost-effective.

We use a stylised valuation:

$$V_{\text{shift}} = N_{\text{risk}} \times R_{\text{reduction}} \times (C_{\text{acute}} - C_{\text{monitor}}), \tag{2}$$

where N_{risk} is the high-risk population under monitoring, $R_{\text{reduction}}$ is the reduction in costly acute events (or progression), and $(C_{\text{acute}} - C_{\text{monitor}})$ is the cost delta between acute episodes and monitored/prevented pathways.

A conservative illustrative scenario (limited to a subset of chronic disease pathways) yields **on the order of €10–30 billion per year**, with a **central** placeholder of **€18 billion/year**. This figure is *highly sensitive* to calibration, adherence, and downstream service design; it should be read as an opportunity bound rather than a forecast.

Mechanism 2: "Airbus for AI": Export and Licensing of Clinical Intelligence If the EU develops the most generalisable and least biased health foundation model — because it is trained on a diverse, multi-system population under strong governance — it could become an exportable strategic asset. Export routes include licensing of model weights within trusted environments, provision of clinical embeddings, and regulated decision-support modules aligned to local pathways.

Rather than relying on a single market-size forecast, we parameterise export value as a function of achievable global share and pricing power. Under a central scenario (modest global share; premium pricing for EU-grade assurance), we use **€12 billion/year** as an illustrative magnitude, with a broad uncertainty range (€5–20 billion/year).

2.3 Summary of calculations

Table **??** consolidates the individual sub-calculations from Layer 1 (efficiency) and Layer 2 (strategic upside) into a single, auditable view. Using the **central** assumptions already stated in the text, the implied annual value at maturity is:

$$V_{\text{total}} = V_A + V_B + V_C + V_{\text{shift}} + V_{\text{export}} = 25.6 + 25 + 10 + 18 + 12 \approx \textbf{90.6 billion €/year}.$$

Table 1: Summary of valuation results (headline numbers, €bn/year).

Value stream (headline)	Low	**Central**	High
Layer 1: Efficiency			
Stream A: Healthcare delivery (efficiency & quality)	10.3	**25.6**	49.5
Stream B: Data reuse & analytics productivity	15.0	**25.0**	35.0
Stream C: Pharma & MedTech R&D acceleration	6.0	**10.0**	15.0
Layer 1 subtotal	31.3	**60.6**	99.5
Layer 2: Strategic upside			
Mechanism 1: "Shift Left" to prevention & monitoring	10.0	**18.0**	30.0
Mechanism 2: "Airbus for AI" export/licensing of clinical intelligence	5.0	**12.0**	20.0
Layer 2 subtotal	15.0	**30.0**	50.0
Grand total (annual value at maturity)	46.3	**90.6**	149.5

Notes: Values are €bn/year. "Low" and "High" reflect the parameter ranges stated (or implied) for each stream and are intended as an illustrative envelope, not a forecast. Totals may differ slightly due to rounding.

To make uncertainty explicit (without changing the core narrative), we also report a simple **low–high** band by combining the parameter ranges already given (or implied) for each stream. The low and high columns are an illustrative envelope constructed from the ranges already given for each stream; they are not the output of a full probabilistic model. Their purpose is to make explicit that (i) the result is not a knife-edge estimate, and (ii) the uncertainty is concentrated in Layer 2, where value depends on calibration, adoption, and governance design rather than on pure technical feasibility.

Table 1 provides a compact audit trail from the paper's mechanism-based decomposition to the headline claim that the EHDS, when coupled to EU-grade foundation models, constitutes a strategic asset on the order of **€90 billion per year** at maturity. The table is intentionally sparse: it reports only the *headline* low/central/high values for each value stream, with each stream's underlying parameterisation already defined in the preceding subsections. This separation keeps the narrative readable while preserving traceability.

The **central** column is the internal consistency check: it shows that the paper's stated central assumptions sum to **90.6** €bn/year (rounded to "€90 bn/year" in the narrative). The **low** and **high** columns are an illustrative envelope constructed from the ranges already given for each stream; they are not the output of a full probabilistic model. Their purpose is to make explicit that (i) the result is not a knife-edge estimate, and (ii) the uncertainty is concentrated in Layer 2, where value depends on calibration, adoption, and governance design rather than on pure technical feasibility.

What drives the central estimate? Three observations matter.

First, the headline figure is not dominated by a single speculative claim. In the **central** case, Layer 1 contributes **60.6** €bn/year, i.e. roughly two-thirds of the total, coming from (A) healthcare delivery productivity and quality improvements, (B) higher productivity in data reuse and analytics, and (C) faster R&D in pharma/medtech where rich longitudinal RWE is binding. These are mechanisms with clear operational analogues (automation, decision support, cohorting, trial design) and direct links to spend baselines.

Second, Layer 2 adds **30.0** €bn/year in the central case, but with a wider uncertainty band. This is exactly the expected pattern if the EHDS is to be interpreted as a *strategic* asset: strategic value is more path-dependent and contingent on implementation quality. Notably, "Shift Left" is the most sensitive component because preventive pathways are economically viable only when prediction is well-calibrated enough to avoid costly false-positive cascades.

Third, the combined result is best interpreted as a *maturity-level* annual run-rate rather than a short-term budget claim. The EHDS is an enabling substrate: its value accrues as adoption rises, as interoperable access becomes routine, and as foundation-model capabilities translate into regulated clinical products and validated workflows. The table, therefore, motivates the policy conclusion of the paper: the EHDS should not be governed and funded as mere "digital plumbing." Under plausible assumptions, it underwrites a European clinical intelligence layer whose economic scale is comparable to major industrial-policy programmes, and whose opportunity cost of delay is correspondingly large.

2.4 Technical Ten-year Net Present Value (NPV) estimation

The headline valuation in Table 1 is an *annual maturity run-rate* (in €bn/year). To translate this into a ten-year strategic business case comparable to standard policy appraisals, we compute a ten-year net present value (NPV) of the EHDS+FM value stack.

NPV definition. We model the EHDS as an enabling substrate whose benefits ramp with adoption, institutional learning, and productisation/validation of foundation-model-enabled workflows. Let $t =$

$1, \ldots, T$ index years after the start of implementation (we use $T = 10$). The present value of net benefits is:

$$\mathrm{NPV}(r) \;=\; \sum_{t=1}^{T} \frac{V(t) - C(t)}{(1+r)^t}, \tag{3}$$

where $V(t)$ is the gross annual value created (across Layer 1 and Layer 2) and $C(t)$ denotes implementation and operating costs (including interoperability roll-out, secure processing environments, assurance, and model lifecycle costs). In the main text we report *gross* values; here we show the gross NPV and note how costs can be incorporated.

Adoption ramp. We express annual value as a maturity run-rate scaled by an adoption factor:

$$V(t) \;=\; V_{\text{mature}} \times a(t), \tag{4}$$

where V_{mature} is the annual value at maturity (central estimate: $V_{\text{mature}} = 90.6$ €bn/year) and $a(t) \in [0, 1]$ captures the fraction of that value realised in year t.

Because institutional adoption rarely jumps instantly to full utilisation, we use a conservative linear ramp as the default (an S-curve would typically be more realistic, but yields similar decade-level magnitudes under comparable mid-point assumptions):

$$a(t) \;=\; a_0 + (1 - a_0)\frac{t-1}{T-1}, \tag{5}$$

with a_0 the first-year realisation fraction. The central case assumes $a_0 = 0.10$ (10% in year 1, reaching 100% by year 10).

Discount rate. We use a real social discount rate r in the range 3–5% and report a central value of $r = 4\%$ for comparability with common public appraisal practice. Results are not knife-edge to this choice because the majority of value accrues in the later years but remains within a single decade.

Central gross NPV. Under the central maturity run-rate (90.6 €bn/year), the linear ramp from 10% to 100% over ten years, and $r = 4\%$, the gross NPV is:

$$\mathrm{NPV}_{\text{gross}} \;\approx\; \sum_{t=1}^{10} \frac{90.6 \times a(t)}{(1.04)^t} \;\approx\; \textbf{€380 billion}. \tag{6}$$

This aligns with the narrative claim of a ten-year NPV on the order of **€350–400 billion** under conservative adoption dynamics.

Incorporating costs (net NPV). If desired, the same framework can compute a net NPV by specifying $C(t)$. Costs can be decomposed into (i) fixed set-up and transition costs (e.g., national connector upgrades, semantic mapping, organisational change), and (ii) recurring operating costs (secure processing environments, assurance, cybersecurity, model maintenance, monitoring, and evaluation):

$$C(t) \;=\; C_{\text{setup}}(t) + C_{\text{run}}(t). \tag{7}$$

In many infrastructure programmes, C_{setup} is front-loaded while $V(t)$ ramps; consequently, the *timing* of costs matters more than the exact accounting category. The key point for this paper is that, because the gross annual value at maturity is in the tens of billions, the decade-level gross NPV remains very large under a wide range of plausible cost profiles.

Sensitivity (how the NPV moves). Two parameters dominate the decade NPV: (i) the maturity run-rate V_{mature} (driven by the Layer 1 and Layer 2 mechanism magnitudes), and (ii) the ramp $a(t)$ (driven by implementation quality, trust, and the operational usability of HealthData@EU). Changes in r within 3–5% shift the NPV but do not change the order-of-magnitude conclusion. Put simply: the strategic question is less "what is the perfect discount rate?" and more "can EHDS governance deliver a high f_{usable} so that adoption and realised value ramp reliably?" This implies that the sociology of implementing and delivering the value of the EHDS is a key driver.

3 Discussion

The Commission's baseline framing of the EHDS produces a sensible infrastructure business case: benefits above €11 billion over ten years at relatively modest incremental cost [1]. However, the emergence of foundation models changes what the EHDS can be: not only a mechanism for cross-border data portability and secondary-use permits, but a potential strategic substrate for EU-grade clinical foundation models that may unlock value at the scale of tens of billions per year.

The core contribution of this paper is not a point estimate. It is the decomposition of value into mechanisms and the identification of which mechanisms are contingent on *scale* and *governance-backed usability*. If EHDS implementation achieves interoperable, secure, high-quality access in practice—as anticipated by both Commission design and TEHDAS operational guidance [3, 12, 13]—then the strategic upside could plausibly be an order of magnitude larger than the "plumbing" business case alone.

3.1 Strategic Implications for Stakeholders

The re-valuation of the EHDS from an €11 billion infrastructure project to a €90 billion strategic asset fundamentally alters the calculus for key stakeholders. The implications of a continental-scale data substrate extend far beyond IT departments, reshaping the landscape for investors, policymakers, and healthcare providers alike.

For the Healthcare Sector: Solving the Productivity Crisis For health systems facing demographic headwinds, the EHDS represents a survival mechanism rather than a technological upgrade. European health systems face a projected shortage of 4.1 million workers by 2030. Foundation Models offer the only scalable means to close this gap. By automating 30–50% of cognitive administrative tasks, such as coding, documentation, and basic triage, the EHDS effectively "manufactures" clinical capacity without adding headcount.

Beyond productivity, the EHDS enables the financial transition from cure to prevention. As demonstrated in our "Shift Left" analysis, the financial viability of healthcare systems depends on moving from expensive acute care to cheap preventive monitoring—a transition that is mathematically impossible without the low false-positive rates enabled by EHDS-scale data. Hospital executives must therefore view the EHDS as the prerequisite for value-based care contracts. Ultimately, this leads to the democratisation of expertise, where high-fidelity Foundation Models act as a "Sidekick" for generalist clinicians, elevating their capabilities to near-specialist levels and improving equitable access across the continent.

For Policy and Politics: From Regulation to Industrial Strategy For policymakers, this analysis demonstrates that the EHDS is not merely a compliance exercise but a cornerstone of industrial strategy and national security. The central issue is data sovereignty. If the EU fails to build its own Health Foundation Models, it will be forced to import them from the US or China, creating a critical dependency on foreign algorithms for healthcare delivery. The EHDS is the only mechanism to ensure "Algorithmic Sovereignty," guaranteeing that European values—privacy, equity, and non-discrimination—are embedded in the weights of the models used in European hospitals.

Furthermore, the economic argument reframes the political narrative from "Privacy vs. Innovation" to "Privacy *and* Protection." Our analysis highlights that the cost of non-Europe—delaying the EHDS—is roughly €60 billion annually in unrealized value. The primary political risk is no longer data misuse, but the *non-use* of data leading to preventable mortality and economic stagnation. In this light, the EHDS should be viewed as the "Airbus of Health." Just as the EU pooled resources to compete in aerospace, the EHDS provides the substrate for a vertically integrated AI stack: European Data, European Compute, and European Models.

For Investors and Venture Capital: A New Asset Class For the investment community, the EHDS effectively resolves the historic structural flaw of European digital health: the fragmentation discount. Historically, European startups have faced 27 distinct regulatory and data regimes, which severely limited their growth potential compared to US peers, who can scale instantly across a single 330-million-person market. By creating a unified addressable market for data access, the EHDS allows ventures to build native, compliant architectures that face lower customer acquisition costs and faster validation cycles. We predict a significant valuation premium for startups that can demonstrate "EHDS-readiness," as they possess a credible path to continental scale that legacy national champions lack.

This shift also demands a re-evaluation of where value accrues in the stack. While the initial wave of value will naturally flow to the infrastructure layer—secure processing environments, privacy-preserving computation, and harmonisation middleware—the long-term alpha lies in the "Model Layer." Vertical AI companies building specialised Foundation Models, such as for oncology or neurology, trained on aggregate EU data, will be the primary beneficiaries. Consequently, investors should prioritise ventures leveraging the "Scale as a Moat" thesis; a model trained on the EHDS corpus holds a defensible data advantage against US competitors, who lack diversity, and Asian competitors, who may lack genetic applicability to Western populations.

Broader Horizons: Resilience and the *In Silico* Shift Beyond the direct economic valuation, the deployment of Foundation Models on EHDS-scale data triggers two structural shifts that fundamentally alter the European Union's risk profile and scientific capacity. While difficult to value using standard cash-flow models, these impacts represent the "insurance value" of the system.

Currently, bio-surveillance is largely retrospective; health systems react to confirmed diagnoses. The EHDS transforms this paradigm by converting the clinical records of 450 million citizens into a real-time, federated "biosensor."

A Foundation Model trained on longitudinal population health is capable of detecting anomalous symptom clusters—such as a statistically improbable spike in respiratory failure or cryptic sepsis in a specific geographic region—weeks before a novel pathogen is biologically sequenced or clinically defined. In this context, the EHDS functions as a critical national security infrastructure. Given that the economic cost of the COVID-19 pandemic to the EU was measured in trillions of euros, an early-warning system that reduces the future containment timeline by even 10% offers an "insurance dividend" that arguably exceeds the system's total efficiency gains.

The Industrialisation of Biology (*In Silico* Revolution) Our analysis of Pharma R&D (Stream C) focused on efficiency. However, the ultimate promise of the EHDS is the transition from *in vivo* experimentation to *in silico* simulation.

The volume of EHDS data allows for the creation of high-fidelity "Digital Twins" of human physiology. This enables the use of Synthetic Control Arms in clinical trials, where new therapies are tested not against a recruited placebo group, but against a statistically perfect digital representation of the target population. This removes the primary bottleneck of scientific progress—patient recruitment—and addresses the ethical dilemma of placebo

3.2 Conclusion

Our AI-aware rigorous re-estimation suggests that the previous top-down analysis captured only roughly 65% of the potential value. The remaining 35% depends entirely on the **scale** of the data: only a continental-scale dataset can drive the accuracy required for prevention and the performance required for export.

The real business case for the European Health Data Space is not a single-digit-billion-dollar savings from nudging doctors to use interoperable software. It is the possibility of creating a **€90 billion-per-year strategic asset** that shifts healthcare from cure to prevention and positions Europe as the global exporter of clinical intelligence.

References

[1] European Commission. *Commission Staff Working Document: Executive Summary of the Impact Assessment Report Accompanying the Proposal for a Regulation on the European Health Data Space*. SWD(2022) 132 final, Strasbourg, 3 May 2022. Available at the European Commission health portal.

[2] European Commission. *Commission Staff Working Document: Impact Assessment Report Accompanying the Proposal for a Regulation on the European Health Data Space*. SWD(2022) 131 final, Strasbourg, 3 May 2022. Available at the European Commission health portal.

[3] European Commission. *A European Health Data Space: harnessing the power of health data for people, patients and innovation*. COM(2022) 196 final, Strasbourg, 3 May 2022.

[4] Eurostat. *Healthcare expenditure statistics (Statistics Explained)*. Accessed 2026.

[5] Eurostat. *EU population increases for the 4th consecutive year (Eurostat news)*. Published 11 July 2025.

[6] MedTech Europe and Deloitte. *The socio-economic impact of AI in healthcare: Addressing barriers to adoption for new healthcare technologies in Europe*. October 2020.

[7] McKinsey & Company. *Generative AI in the pharmaceutical industry: moving from hype to reality*. 9 January 2024.

[8] Kaplan, J., McCandlish, S., Henighan, T., et al. Scaling Laws for Neural Language Models. *arXiv:2001.08361*, 2020.

[9] Hoffmann, J., Borgeaud, S., Mensch, A., et al. Training Compute-Optimal Large Language Models. *arXiv:2203.15556*, 2022.

[10] Wei, J., et al. Emergent Abilities of Large Language Models. *arXiv:2206.07682*, 2022.

[11] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B. High-Resolution Image Synthesis with Latent Diffusion Models. In *Proceedings of CVPR*, 2022.

[12] TEHDAS Joint Action. *Recommendations on a Data Quality Framework for the EHDS*. September 2023.

[13] TEHDAS Joint Action. *Report on architecture and infrastructure options to support EHDS services for secondary use of data*. March 2023.