White Paper

Secondary Use of Health Data, from Legislation to Implementation: Establishing Trust Through De-identification in the European Health Data Space (EHDS2)

CHRIS VAN BRONCKHORST, Senior IT and Digital Health Professional, EU Institutions, IQVIA
JONATHAN GREEN, Director, IQVIA's Privacy Analytics
LUK ARBUCKLE, Global AI Practice Leader and Chief Methodologist, IQVIA's Privacy Analytics

Table of contents

Acknowledgements

Olivia Lounsbury (University of Oxford), Jordan Collins (IQVIA's Privacy Analytics) Donna Franklin (IQVIA's Privacy Analytics), Susanna Flaherty (IQVIA), and George Garrard (IQVIA) supported the development, ideation, and revision of this paper.

Introduction

The EHDS2 Regulation marks a milestone in the secure reuse of health data within the EU, with trust and privacy at its core. This paper highlights the importance of a centralized de-identification function under the supervision of National Health Data Access Bodies (HDABs). Such an approach promotes consistency, transparency, and interoperability, while preventing fragmentation and the risk of re-identification. It is a strategic investment in the legitimacy and scalability of European health data policy.

This paper is clearly structured into six parts that together bridge the gap between legislation and implementation within the EHDS2. Each section addresses a key aspect of building a trustworthy and effective European Health Data Space. The guide below helps you quickly navigate the content.

1. **Executive summary** — Summarizes the core message and recommends a centralized de-identification function.
2. **The legislation and de-identification** — Explains the legal framework and the role of privacy protection through de-identification.
3. **Implementation challenges** — Outlines practical obstacles such as capacity gaps, unclear responsibilities, and fragmentation.
4. **Proposed solution** — Presents a centralized service to perform de-identification on behalf of all involved parties.
5. **Key benefits** — Describes five advantages of a centralized approach: consistency, transparency, contextual adaptation, integration, and scalability.
6. **Earning and sustaining trust** — Shows how a centralized model supports lasting public trust and successful EHDS2 implementation.

Part 1 — Executive summary

This paper is intended for Health Data Access Bodies (HDABs), ministries of health, and organisations responsible for implementing the EHDS2 secondary use framework. It is designed to support policy, architectural, and governance decisions on how de-identification should be operationalised at national or regional level. The paper is written for a point in time where Member States are moving from legislative transposition to system design, procurement, and early operational delivery, and where foundational design choices will have long-term implications for interoperability, scalability, efficacy, and public trust.

The European Health Data Space (EHDS2) Regulation (EU) 2025/327 represents a transformative step toward enabling the secure and trustworthy secondary use of health data across the European Union. As Member States move from legislative transposition to system design, procurement, and operational delivery, a central challenge emerges; **how to operationalise de-identification in a way that is legally robust, technically effective, and capable of sustaining public trust at scale**.

At the heart of EHDS2 lies a foundational principle: trust. Citizens' willingness to allow the reuse of their health data for research, innovation, and policymaking depends on credible assurances that privacy is protected through transparent, consistent, and accountable data governance. De-identification, through anonymisation or pseudonymisation, is the Regulation's primary technical mechanism for delivering that assurance.

EHDS2 establishes a layered legal approach. Where anonymised data is sufficient, it should be used by default. Where pseudonymised data is requested, Article 68(1)(c) requires HDABs to assess whether such use is sufficiently justified. This approach aligns closely with the GDPR principle of Data Protection by Design (Article 25 GDPR). However, translating this legal intent into operational practice across a highly diverse ecosystem of Health Data Holders (HDHs) and a variety of health data categories presents significant challenges.

Key implementation barriers include capacity and expertise gaps among HDABs and HDHs, particularly smaller providers such as general practitioners, pharmacies, and mental health services; ambiguity over roles and responsibilities in the de-identification process; and the risk of fragmented national practices that could undermine both privacy protection and data interoperability. Without a coherent operational model, Member States face the twin risks of inconsistent privacy safeguards and reduced data utility.

To address these risks, this paper proposes the establishment of a centralised de-identification function at national or regional level, operating under the mandate and oversight of the HDAB. **Rather than distributing complex and high-risk de-identification tasks across hundreds of individual HDHs, this function would act as a shared, expert capability, executing de-identification on behalf of HDHs and HDABs as part of the EHDS2 access workflow**.

A centralised de-identification function delivers five core benefits:

i. **Consistency, data linkage, and interoperability** — uniform de-identification rules applied over time and across data sources reduce fragmentation, enable lawful data linkage within and across countries, and support cross-border interoperability, an essential prerequisite for a functioning European Health Data Space.

ii. **Transparency and accountability** — centralised execution enables comprehensive, auditable logging of all de-identification activities, supporting regulatory oversight, public reporting, and citizens' rights to understand how their data is protected.

iii. **Use-case-specific de-identification** — through the development of a de-identification playbook defining transformation rules and controls, aligned with the data categories defined in Article 51 and the intended purpose of use, the central function enables tailored approaches that reflect both the sensitivity of each data asset and the analytical needs of the use case.

iv. **Integration with the Secure Processing Environment (SPE)** — de-identification is performed within, or as an integral component of, the SPE or a dedicated Privacy Enhancing Technology capability supporting it, ensuring that identifiable data never leaves a controlled environment and reducing re-identification risk.

v. **Efficiency and scalability** — pooling expertise, technology, and processes reduces duplication, accelerates turnaround times, and enables early feasibility assessments during the permit process, aligning privacy protection with realistic research expectations.

By embedding de-identification within a centralised function that meets high standards of consistency, transparency, contextual adaptation, and technical integration, Member States can move from legal compliance to operational credibility. This approach is not merely a technical optimisation; it is a strategic investment in the long-term legitimacy, interoperability, and public trustworthiness of the EHDS2 ecosystem.

This paper therefore recommends that Member States implement de-identification as a centralised function, operating under the mandate and oversight of the HDAB, rather than distributing responsibility across individual health data holders.

Part 2 — The legislation and de-identification

Trust through data privacy

EHDS is built on the principle that public trust is a prerequisite for the secondary use of health data. The Regulation's core privacy safeguard is stringent de-identification of personal health data (anonymization or pseudonymization) before any secondary processing. In legal terms, this means that only data which is not directly identifiable may be used for research, policy, innovation, and other permitted purposes. Article 66 of the EHDS2 mandates that HDABs make data available "in anonymised form — or in pseudonymised form if the user justifies that the research purpose requires it". By default, fully anonymised data (no individual can be identified, considering all means "reasonably likely" to be used) is preferred. Pseudonymised data (where identifying details are replaced by codes held separately) can be provided only when necessary and under strict controls. This hierarchical approach reflects the Regulation's GDPR-based ethos: maximize privacy by design and only allow identifiable elements if proportionate and essential.

Responsibilities and process — from source to secure environment

The EHDS2 legal framework defines clear roles in the data supply chain, while expecting Member States to flesh out operational details. HDHs — broadly defined in Article 2(1)(t) to include any entity that controls electronic health data — range from large hospitals and national registries to primary care practices, pharmacies, and mental health clinics in the community. These HDHs have a legal duty to provide requested datasets to the HDAB for approved secondary uses (Article 51). They must register descriptions of their datasets in a national catalog to facilitate discovery (Article 60(3)). Crucially, before sharing data, HDHs are expected to respect patients' choices by excluding individuals who have exercised their right to opt out of secondary use (Article 71). The Regulation allows each country to decide whether the responsibility for enforcing opt-outs lies with the HDH, the HDAB, or a Trusted health data holder (THDH) — a HDH formally recognized for their additional governance, security, and performance capabilities. However, in all cases this must be clearly assigned in national law. This ensures that citizens' explicit right to object to reuse of their health data is operationalized as a legal safeguard.

Following approval of a data permit, the HDAB coordinates the flow of data from the relevant HDHs. HDABs are the central governing entities established or designated by each Member State (Article 55) to oversee secondary use requests. They are the primary guarantors of data privacy and security in the EHDS2 system. Under Article 57, the HDAB must ensure that only the minimum necessary data is shared and that all required de-identification measures are applied. The law specifies that this de-identification should occur "as early as possible in the chain" of data provision (Recital 72). In practical terms, Recital 72 suggests that anonymization ideally happens at the source (e.g., at the GP practice or pharmacy before data leaves its system) whenever feasible. However, recognizing that not all HDHs have the capability to do this properly, the Regulation permits that either the HDH or the HDAB may perform the necessary anonymisation or pseudonymisation. The HDAB therefore carries ultimate responsibility for ensuring de-identification, even if the task is executed by others. Notably, accountability cannot be outsourced: if a HDH or a specialized third party conducts the de-identification, the HDAB must verify and bear responsibility for its efficacy.

After the data is de-identified, the HDAB provides access to it through a Secure Processing Environment (SPE) (Article 73). The SPE is a controlled, non-public data sandbox that meets high security and privacy standards (defined via the Data Governance Act and further detailed by EHDS2 implementing acts). Health Data Users (HDU, e.g., researchers) access and analyze the permitted dataset within the SPE — they cannot download or remove raw data files. Only aggregated results or outputs that have passed disclosure control may be exported from the SPE. This mechanism is designed to ensure that even pseudonymised data

does not leave the SPE. According to Article 57 and 73, the HDAB must monitor HDU's activities in the SPE and audit compliance with the permit conditions (e.g., ensuring the HDU does not attempt re-identification or use data beyond the allowed purpose). Notably, the law frames the HDAB (or a designated THDH operating its own SPE) as a joint controller for data preparation and de-identification steps, even as it is simultaneously considered a "processor" vis-à-vis the HDU for providing the technical platform. This complex legal characterization underscores that HDABs are at the fulcrum of both enabling data use and enforcing data protection.

Citizens' rights and transparency

In addition to technical measures, EHDS2 fortifies trust through individual rights and transparency obligations. Every person in the EU has the right to opt out of the secondary use of their health data (Article 71), responding directly to public concerns about involuntary data sharing. This opt-out must be easy to exercise, for example via a national digital portal or during an interaction with a healthcare provider. Crucially, once an opt-out is registered, it applies to the individual across the Member State, rather than to a specific provider or care setting, and is binding on all HDHs and HDABs within that jurisdiction.

Transparency around data use is reinforced through mandatory public disclosure. Article 58 requires HDABs to publish information on all data permits they issue, including the purpose of use, the categories of data involved, and the application and results, with many countries expected to maintain public online registers of approved projects. Citizens can consult these registers to understand how health data is being used for research or policy purposes. In addition, upon request, individuals should be informed whether their own data was included in any released dataset, a logical extension of transparency that is likely to be implemented through national or regional digital health portals. Together, these measures echo the UK National Data Guardian's recommendations following the care.data experience: make data use visible and give people a meaningful choice in order to build public confidence.

In summary, the EHDS2 legislation establishes a comprehensive policy framework where de-identification is integral to data governance. HDABs are charged with operationalizing this framework: they must coordinate de-identification and secure access, working with a wide array of HDHs including small-scale providers like general practitioners, community pharmacies, and mental health services. The legal texts provide guiding principles (e.g., anonymize early, share only what is needed, documenting all actions) and allocate accountability, but leave implementation details to Member States' discretion. This combination of binding rules and flexible execution aims to ensure robust privacy protection while accommodating diverse health system structures across Europe.

The next section examines the challenges that arise in turning these legal mandates into practical, effective operations on the ground.

Part 3 — Implementation challenges in achieving effective de-identification

Implementing EHDS2's de-identification requirements will be complex. While the Regulation sets out a high-level vision, several challenges must be addressed to bridge the gap between legislation and real-world practice.

The challenges

CAPACITY AND EXPERTISE CONSTRAINTS

A significant challenge is the shortage of specialized expertise among many HDHs and even HDABs. De-identification of health data — especially free-text clinical notes, detailed medical records, or genomic datasets — is a technically demanding process requiring knowledge of privacy techniques (e.g., k-anonymity, noise addition, suppression, generalization) and context-specific risks. Larger national institutions or university hospitals may have data protection officers or IT departments familiar with such techniques, but small-scale HDHs like local customary practices, individual pharmacies, or mental health clinics typically do not have in-house data scientists or privacy engineers. These front-line care providers are suddenly cast in the role of data providers under EHDS2, expected to handle tasks like pseudonymizing patient identifiers or aggregating rare data points before transmitting data to an HDAB. Without additional support, there is a risk of uneven de-identification quality — some providers might inadvertently fail to remove all identifiers (jeopardizing privacy), while others might over-sanitize data (diminishing its utility). Member States will need to invest in training programs, technical tools, or shared services to uplift the capabilities of these diverse HDHs. The alternative is an uncoordinated patchwork of practices that could undermine both privacy and the scientific value of the data.

AMBIGUITIES IN ROLES AND PROCEDURES

The EHDS2 text leaves certain operational questions open, requiring clarification through implementing acts or national implementation choices. A prime example is the question of "who exactly performs the de-identification, and when." The law establishes HDABs as accountable for ensuring it is done, but as noted, it permits either the HDH or HDAB to carry out the actual work. This flexibility acknowledges varying national infrastructures but creates potential ambiguity. Without clear guidelines, one country's health data might be consistently anonymized at source by clinics and pharmacies, while another country might funnel identifiable data to a central hub for anonymization. Such differences could complicate cross-border data sharing if, for instance, datasets from two countries have been de-identified using different standards or at various stages. Consistent interpretation of "de-identification as early as possible" is needed. National laws or guidance will have to specify, for instance, how a general practitioner should transmit data to the HDAB (e.g., via a secure platform) and at what point identifying information is stripped. Any lack of procedural clarity can slow implementation and erode accountability, as no stakeholder wants to act beyond their mandate when handling sensitive patient data.

MAINTAINING FIT-FOR-PURPOSE DATA

A central implementation challenge is achieving the right balance between privacy protection and data usefulness. The EHDS2 principle of data minimization means only the data strictly necessary for the approved purpose should be shared. In practice, this requires careful judgement. Over-aggressive anonymization could render data not fit for purpose - for example, if a study on medication safety in elderly patients only receives age ranges like "18-65" and "65+" instead of more granular age brackets, critical insights might be lost. On the other hand, if data are too detailed (e.g., exact birth dates, full postal codes in rural areas), individuals could be re-identified, violating privacy. The legislation implicitly acknowledges this tension: if an applicant can convincingly justify that anonymized (aggregated) data will not answer their research question, the HDAB may allow pseudonymised person-level data but under very strict conditions (within the SPE, with no external data linking unless approved, etc.). Implementers must develop clear criteria for these decisions. A use-case-

based approach is needed; for each type of analysis, define what degree of detail is permissible. For instance, public health statistical reporting might only need de-identified aggregates, whereas a machine learning research project might require patient-level longitudinal data (pseudonymized but rich). This balancing act is challenging because it demands interdisciplinary expertise: understanding the health data, legal risks, privacy science, and the research data requirements. Without careful calibration, EHDS2 could face two failure modes: researchers receiving datasets so scrubbed of information that they abandon using them, or worse, a privacy incident where someone's identity is exposed from a supposedly de-identified dataset. Both outcomes would damage the credibility of the system. Robust guidelines and perhaps simulation exercises (testing whether results still hold after proposed anonymization steps) will be important to ensure data remains useful while adhering to privacy norms.

INCONSISTENCY AND FRAGMENTATION RISKS

The Regulation sets baseline requirements but leaves room for divergent national implementations, which could lead to inconsistencies. For example, one Member State might adopt a very strict anonymization standard (interpreting "anonymised" to mean nearly zero re-identification risk under any scenario), while another takes a more contextual approach (allowing data to be considered anonymized if re-identification would require unreasonable effort, in line with evolving GDPR interpretation). Differences could also arise in technological standards for the SPEs or in how HDABs audit compliance. Such fragmentation would undermine the seamless European data space ideal and could reduce mutual trust between countries. The EHDS2 governance structure — including an EHDS Board at EU level — is intended to foster harmonization by developing common guidelines and facilitating exchange of best practices and propagate Europe-wide standards such as the January 2025 EDPB Guidelines on Pseudonymisation, which provide detailed best practices relevant to EHDS2.

TRANSPARENCY AND ACCOUNTABILITY MECHANISMS

Implementing EHDS2 also involves setting up new accountability tools. HDABs must log and audit every action taken on health data — from the initial extraction by a HDH (e.g., a mental health center assembling records) to the transformations performed (masking names, aggregating sensitive attributes, etc.) to user interactions in the SPE. This end-to-end traceability is vital for compliance and public accountability, but building such an audit infrastructure is non-trivial. Many health systems currently lack integrated logging across institutions. Under EHDS2, if a data subject asks "Who accessed my data and what was done to protect it?", the system should be able to provide a clear answer, e.g.: "Your GP provided a pseudonymised record to HDAB on Date X; identifiers were replaced or masked; your data was included in analysis Y under permit Z; results were published at [link]." Producing this kind of trace requires interoperable record-keeping across multiple entities, potentially using unique permit IDs and dataset IDs that all parties reference. Policymakers will need to ensure that national implementations include IT solutions and legal mandates for such record-keeping. Additionally, the success of EHDS2 will be judged by the public through evidence of its trustworthiness. Annual reports (as required by Article 59) should not be perfunctory but should meaningfully disclose how privacy was safeguarded. Clear communication in these reports can reinforce trust that the system is being used responsibly. Without visible accountability, even a well-functioning system might struggle with public perception.

Past and current experiences

Historical lessons underscore these challenges. The failure of NHS care data in England (2013–2016) is often cited by policymakers as a warning. The care data program aimed to centralize patient records for research and care improvement, but it faltered due to insufficient transparency, an unclear opt-out process, and public concern over data privacy and potential misuse. A government-commissioned review concluded that a robust opt-out and better de-identification treatment were necessary to regain public confidence. On the other hand, positive examples exist. Finland's Findata system (a national one-stop data permit authority launched in 2019) has demonstrated that with the right infrastructure and expertise, a centralized approach can handle large volumes of requests while maintaining privacy. In 2025 alone, Findata received hundreds of applications and issued 367 secondary-use data permits/decisions (94% approvals) under Finland's secondary use of health and social data law. Also, 85% of the applications were processed in less than three months. The Finnish experience shows that a central agency can effectively compile data from dozens of controllers (over 60 in one case) and apply consistent de-identification before release. Finland invested in multiple accredited secure data environments (ten nationwide) with about 5,000 authorized users, supporting more than 1,100 research projects, all coordinated by Findata. This level of organization and capacity did not emerge overnight; it required political commitment, resources, and clear protocols. Other EU countries will need similar investments to meet EHDS2's promise.

In light of these challenges, strategic solutions are required to move "from legislation to implementation" without eroding the trust that the law seeks to build. In the concluding section, we present a policy vision from IQVIA's Privacy Analytics' (PA) for such a solution: a centralized de-identification function that could help Member States ensure consistent and high-quality de-identification, thereby strengthening the foundations of trust in the EHDS2 ecosystem.

Part 4 — Privacy Analytics' proposed solution: A centralized de-identification function to build trust

Addressing the challenges by implementing a centralized de-identification process

To address the above challenges and based on Privacy Analytics' (PA) experience and knowledge of various EU privacy functions, PA proposes that Member States establish a central de-Identification service as part of their EHDS2 implementation. The concept is to create one specialized unit or platform at the national (or major regional) level responsible for performing data anonymization and pseudonymization tasks on behalf of all HDHs and HDABs in that jurisdiction. This service would operate under the mandate and oversight of the coordinating HDAB, but it would concentrate the technical execution of de-identification in a single, expert-driven process. Instead of hundreds of clinics, pharmacies, and institutions each devising their own de-identification solutions, one trusted service would handle it for all, using state-of-the-art methods and tools.

The challenges outlined in the previous part, ranging from capacity constraints and procedural ambiguity to fragmentation, data usability concerns, and limited traceability, require a coherent and scalable response. This Part 4 presents a policy for a centralised de-identification function that addresses these challenges through five mutually reinforcing principles: consistency/uniformity, transparency, specificity, control, and efficiency/scalability. Together, these elements form a practical and policy-aligned solution to operationalise the EHDS2's privacy requirements while maintaining data utility and public trust.

To illustrate how these policy elements directly respond to the identified challenges, the matrix shown in figure 1 maps each challenge against the corresponding components of the proposed policy. It demonstrates how the centralized approach offers a comprehensive and structured response to the operational complexities of EHDS2 implementation. The remainder of this paper elaborates on each of these policy elements in detail.

Figure 1: challenge/policy matrix

		POLICY ELEMENTS				
		1. **Consistency/ Uniformity**	2. **Transparency**	3. **Use-case specificity**	4. **Integration with SPE**	5. **Efficiency and scalability**
CHALLENGES	1. **Capacity gaps at HDABs and HDHs**	Uniform standards reduce the need for HDHs to develop their own methods, easing pressure on limited internal capacity.			Centralising de-identification within or near the SPE reduces the need for local technical infrastructure.	Centralisation and automation reduce reliance on local expertise and enable economies of scale.
	2. **Ambiguity in roles and responsibilities**	Uniform processes clarify responsibilities across national contexts and institutions.	Centralised logging and reporting make responsibilities visible and verifiable.		Linking de-identification to the SPE makes HDAB accountability explicit.	Central execution simplifies the division of labour between HDABs and HDHs.
	3. **Inconsistency and fragmentation across actors**	A single centralised approach prevents divergent interpretations and implementations across actors.	Central reporting highlights discrepancies and supports harmonisation.	The playbook ensures uniform application per data category and use case.	Integration with the SPE enforces a shared technical and procedural baseline.	Scalable central services allow less-resourced actors to align with a harmonised system.
	4. **Difficulty maintaining data usability**		Transparency about applied techniques helps users interpret and trust the data.	The playbook enables tailored de-identification that preserves utility for each use case.		Early feasibility assessments prevent delivery of over-sanitised or unusable datasets.
	5. **Limited traceability and public accountability**	Uniform and consistent processes enable systematic logging of all actions.	Centralised audit trails and reporting support accountability to citizens and regulators.		Processing within the SPE ensures every step is verifiable and attributable.	Centralisation enables more efficient and complete audit and reporting mechanisms.

Part 5 — Key benefits of a centralised approach

The centralised de-identification function responds directly to the operational challenges outlined in Part 3. This section elaborates on how the five interdependent policy elements together form a coherent and practical solution. Each element is grounded in legal obligations, operational feasibility, and lessons learned from past initiatives.

Consistency and uniformity

Consistency ensures that similar data processing tasks are performed in the same way over time, while uniformity ensures that similar tasks are handled similarly across different actors and contexts at a given moment. Both are essential to ensure legal certainty, interoperability, and trust in the system.

By design, a single service enforces consistent and uniform application of de-identification rules across all datasets and projects. This directly combats fragmentation and ensures that HDUs receive comparable outputs regardless of the data source. For example, if two researchers request similar datasets from different regions — say, prescription data from pharmacies — the central function ensures that age bands, suppression thresholds, and masking techniques are applied identically.

To institutionalise this, the central function maintains a detailed "de-identification library" (or playbook) that enumerates standard measures for each data type and context. This library aligns with the data categories of Article 51 (e.g., EHR data, genetic data, administrative claims) and specifies how to handle each category. For instance, for prescription data from community pharmacies, the library might stipulate removal of patient names and addresses, aggregation of patient age into bands (to prevent identification of especially old or young patients), and generalisation of rare diagnoses or drug combinations. For mental health service records, which may contain extremely sensitive information, the library could call for additional steps like redacting detailed narrative fields or using code frames for diagnostic notes.

These measures are selected based on both legal requirements and empirical risk assessments, aiming for the optimal balance between privacy protection and analytical value. Crucially, the library ensures that if two applicants request similar data in different regions or time periods, they receive comparably de-identified datasets. This not only ensures fairness for HDUs but also facilitates data pooling and supports regulatory oversight. The library is maintained and updated by the central service in consultation with the (coordinating) HDAB and expert bodies, allowing it to evolve in response to new threats, technologies, or legal interpretations.

Transparency

Transparency is a prerequisite for public trust and regulatory accountability. The central function enables full traceability of de-identification activities through systematic logging, audit trails, and documentation. Every transformation, whether suppression, generalisation, or pseudonymisation, is recorded and linked to the specific permit, use case, and HDU.

This traceability allows HDABs to demonstrate compliance with Articles 57 and 59 of the EHDS2 Regulation and supports oversight by data protection authorities. It also enables meaningful public reporting. For example, HDABs could publish annual statistics on the proportion of permits involving anonymised versus pseudonymised data, or on the most frequently applied de-identification techniques. Such reporting reinforces the message that data reuse is governed by robust safeguards.

Moreover, transparency supports internal accountability. If a data subject exercises their right to know how their data was used, the system should be able to provide a clear, verifiable account: when the data was extracted, how it was transformed, and under which permit it was accessed.

Use case specificity

De-identification is not a binary process. The level and method of de-identification must be tailored to the context in which the data will be used. A centralised function enables this by implementing a consistent and structured, use-case-based approach.

The de-identification library plays a central role in this. It defines appropriate techniques for each data type and analytical purpose. For example, in a study on medication adherence using pharmacy dispensing data, the library may prescribe suppression of rare drug combinations and aggregation of age into five-year bands.

This approach ensures that data remains analytically useful while respecting privacy. It also provides clarity to HDABs and applicants during the permit process: both parties can assess in advance whether the requested data can be sufficiently de-identified to meet the research objectives.

Integration with the Secure Processing Environment (SPE)

The SPE is the technical and procedural backbone of EHDS2's privacy framework. Integrating the de-identification process within or directly adjacent to the SPE ensures that identifiable data never leaves a controlled environment. This design minimises the risk of unauthorised access or re-identification and simplifies compliance with data protection requirements.

For example, raw data from multiple HDHs can be securely ingested into the SPE, where the central function applies the approved de-identification protocol. Only the resulting de-identified dataset is made accessible to the HDU, and only within the SPE. This ensures that even pseudonymised data is never exported or linked outside the secure perimeter. It also clarifies the accountability of the HDAB, which oversees both the de-identification process and the user's activity within the SPE.

This model is already operational in Finland, where Findata performs deidentification during data processing before the data is transferred to the Kapseli® SPE. The Finnish experience demonstrates that such integration is both feasible and effective.

Efficiency and scalability

To meet the ambitions of EHDS2, Member States must be able to process large volumes of data requests in a reliable, timely, and cost-effective manner. A centralised de-identification function offers a scalable and future-proof model that supports operational efficiency and policy robustness.

A well-resourced central service enables economies of scale. By pooling expertise, technology, and processes, Member States can invest in high-quality capabilities that are accessible to all participating actors. Rather than requiring each individual HDH, such as a general practitioner or pharmacy, to develop its own de-identification capacity, the central function provides a shared infrastructure that assumes this responsibility. This lowers the threshold for participation in EHDS2, particularly for smaller or less technically equipped organisations.

Automation plays a key role in this model. The central function can deploy advanced tools such as pseudonymisation engines and Natural Language Processing (NLP) to automatically detect and mask personal data in unstructured content, such as free-text fields in clinical records. E.g., Findata processes open-text data using NLP machine learning methods, more specifically an NER (Named Entity Recognition) model (BERT-based). As processes become standardised and optimised, turnaround times for delivering datasets are expected to improve significantly.

An additional benefit is the ability to conduct early feasibility assessments during the permit application process. The central service can evaluate the requested data and research plan to determine whether the desired outputs can be achieved using sufficiently de-identified data. If not, the service can proactively suggest alternatives. For example, if a researcher requests highly granular location data that would be difficult to anonymise, the central function may recommend using a higher level of geographic aggregation. This consultation occurs before the permit is issued, allowing the HDAB to define realistic conditions and ensure that the applicant understands any limitations in advance. This proactive guidance helps prevent situations where HDUs receive over-redacted datasets that fail to meet their analytical needs, scenarios that can lead to frustration and erode trust in the system.

In practice, the central service acts as a privacy advisory partner to both HDABs and applicants, ensuring that each project is scoped from the outset with a realistic balance between privacy protection and research value.

This approach not only increases efficiency but also ensures the scalability of the system. It is essential for the sustainable implementation of EHDS2 across a diverse European landscape with varying levels of capacity and resources.

In summary, PA advocates a centralized approach to de-identification in each member state or even shared across regions for smaller countries, and directly addresses gaps in expertise, consistency, and traceability.

Part 6 — Earning and sustaining trust

Ultimately, the goal of a centralized de-identification function is to operationalize the legal requirements of EHDS2 in a way that is scalable, reliable, and worthy of public trust. Trust is fragile; it must be continuously earned through performance and openness. By investing in a centralised model, authorities signal that they are taking the privacy mandate seriously — not leaving it to chance or uneven capabilities. This model makes it easier to communicate to the public how their data is protected: one can point to the central service's protocol and audits as proof of rigorous safeguards. It also simplifies compliance with the letter and spirit of the law since many obligations (from "anonymize early" to "document everything") can be baked into one coordinated process. As EHDS2 moves from legislation to implementation, the world will be watching how the EU balances innovation with privacy. A well-implemented central de-identification function directly addresses the major concerns (privacy risks, inconsistent practices, capacity shortfalls) that could otherwise derail the initiative. It embodies the idea of Data Protection by Design, as required by GDPR Article 25, on a national scale, ensuring that privacy considerations are embedded at every step of the data journey. By doing so, it lays the groundwork for a trusted European Health Data Space where stakeholders, patients, providers, and health data users, feel confident that the reuse of health data is being conducted safely, lawfully, and for the collective good. This trust, once earned, will enable EHDS2 to achieve its transformative potential for health research and innovation across Europe.

Glossary

Anonymisation

A process by which personal data is irreversibly transformed so that an individual can no longer be identified, taking into account all means reasonably likely to be used. Once anonymised, data falls outside the scope of the GDPR.

Pseudonymisation

The processing of personal data in such a way that individuals cannot be identified without additional information, which is kept separately and subject to technical and organisational safeguards. Pseudonymised data remain personal data under the GDPR.

De-identification

An umbrella term used in this paper to refer to the application of anonymisation or pseudonymisation techniques, together with supporting controls, to reduce the risk of identifying individuals in health data used for secondary purposes.

Health Data Access Body (HDAB)

A public authority designated or established by a Member State under EHDS2 to assess secondary-use applications, issue data permits, and ensure compliance with legal, privacy, and security requirements.

Health Data Holder (HDH)

Any natural or legal person that controls electronic health data and is required under EHDS2 to make such data available for approved secondary uses.

Health Data User (HDU)

An organisation or individual granted a data permit by an HDAB to access health data for an approved secondary use, such as research, innovation, or policymaking.

Secure Processing Environment (SPE)

A controlled technical environment in which health data are made available for secondary use under EHDS2. The SPE restricts access, prevents unauthorised data extraction, and ensures that only approved outputs may leave the environment.

Privacy Enhancing Technologies (PETs)

Technical measures designed to protect personal data while enabling data use, including techniques such as pseudonymisation, secure environments, access controls, and disclosure control.

Data protection by design

A principle under Article 25 of the GDPR requiring data protection safeguards to be embedded into systems and processes from the outset, rather than applied retrospectively.

Privacy Analytics, an IQVIA company

Privacy Analytics is a specialist data privacy and governance team focused on enabling the safe, lawful, and trusted use of data for research, innovation, and public benefit. The team's mission is to help organisations move from privacy principles and regulation to practical, operational solutions that protect individuals while preserving the analytical value of data.

Privacy Analytics works with public authorities, health systems, and data-intensive organisations to design and implement privacy-by-design architectures, with particular emphasis on de-identification, governance models, and secure data access environments. The team's work spans policy advisory, system design, and operational delivery, supporting organisations as they navigate complex regulatory frameworks such as the GDPR and the European Health Data Space.

In the UK, Privacy Analytics has been closely involved in the development and delivery of NHS Privacy Enhancing Technology (PET) capabilities, supporting the operationalisation of de-identification and secure access within NHS data environments. The NHS PET experience provides a practical reference point for EHDS2 implementation, demonstrating how centralised de-identification, strong governance, and secure processing environments can be combined to enable secondary use of health data at scale while maintaining public trust.

Beyond the UK, Privacy Analytics has contributed to VALO 1 and VALO 2, Nordic initiatives exploring technical, governance, and operational capabilities relevant to EHDS implementation. These projects tested approaches to cross-organisational data access, de-identification, and secure processing in environments aligned with EHDS objectives, offering early insights into how EHDS-style architectures can function in practice across different national contexts.

Together, these engagements inform Privacy Analytics' policy perspective in this paper: that centralised, expert-led de-identification functions, embedded within secure processing environments and governed by clear accountability, are critical to translating EHDS2 from legislation into trusted, operational reality.

About the authors

CHRIS VAN BRONCKHORST

Senior IT and Digital Health Professional, EU Institutions, IQVIA

Chris van Bronckhorst is a senior IT and digital health professional with a strong technical foundation spanning software development, data engineering, and enterprise architecture, evolving into advisory and leadership roles in complex digital transformation programmes. His expertise lies at the intersection of health IT, data platforms, and regulatory-driven data sharing.

He has in-depth experience with the secondary use of health data, including real-world data and evidence, interoperability standards, data governance, and privacy-preserving data processing. Chris is closely involved in initiatives aligned with the European Health Data Space (EHDS), with particular focus on federated analytics models that enable cross-border analysis while data remains with local health data holders.

In his current role, he leads technical teams delivering advanced health data services to authorities like the European Medicines Agency (EMA), supporting pharmacovigilance, regulatory analytics, and observational research. His team translates legal, policy, and scientific requirements into robust technical designs, acting as a bridge between data scientists, security specialists, and public authorities to deliver scalable, compliant, and future-proof health data services.

JONATHAN GREEN

Director, IQVIA's Privacy Analytics

Jonathan Green is a Director at IQVIA's Privacy Analytics, Europe. He collaborates actively with Governments and Healthcare Providers across Europe, Middle East and Africa to support the development of their data transformation solutions and readiness for European Health Data Space, enabling them to achieve the careful balance between maintaining public trust and the security of the patient's data, while maximising the full value of the data for its intended use.

As a regulatory lawyer with 30+ years post qualification experience, Jonathan has worked across diverse healthcare settings. Jonathan is currently the Programme Director of IQVIA's Privacy Enhancing Technology collaboration with the UK National Health Service Federated Data Platform and is Data Governance Specialist Adviser to the VALO2 Project in the Nordics, which is demonstrating how privacy-preserving, federated analytics can enable secure, compliant cross-border health data use in line with the European Health Data Space vision.

LUK ARBUCKLE

Global AI Practice Leader and Chief Methodologist, IQVIA's Privacy Analytics

Luk Arbuckle is Global AI Practice Leader and Chief Methodologist within IQVIA Applied AI Science, where he leads the development of defensible, trustworthy AI and data-driven systems for healthcare and life sciences. His work focuses on aligning advanced analytics with regulatory, ethical, and scientific expectations at scale.

Alongside this role, Luk provides methodological leadership within Privacy Analytics, IQVIA's specialist privacy and governance function. He advises public authorities and data-intensive organizations on de-identification, privacy-by-design architectures, and accountable data access models that enable secondary use of health data while sustaining public trust.

Luk brings a cross-disciplinary background spanning engineering, signal and image processing, statistics, geointelligence, privacy-enhancing technologies, and regulatory oversight. His experience includes roles in industry, a data protection authority, and healthcare research, and he regularly contributes to international standards, policy dialogue, and applied guidance on trustworthy AI and data governance.

CONTACT US

IQVIA Solutions Belgium

Da Vincilaan 7 — Davos building

1930 – Zaventem

Chris.vanbronckhorst@iqvia.com

Jonathan.Green@iqvia.com

luk.arbuckle@iqvia.com

iqvia.com