Trusted Data Exchanges Across Regulatory Boundaries: A Research Agenda for Personal Data Vault Interoperability

Marlin Udo Kisia[1], Ruben Verborgh[2], and Beatriz Esteves[1]

[1] IDLab, Dept. Electronics and Information Systems, Ghent University – imec, Belgium
[2] Ghent University, Belgium
{firstname.lastname}@ugent.be

Abstract. As the European Union pursues a single market for data, personal data vaults promise individuals a genuine role in the governance of data concerning them. Yet a single exchange leaving a vault is rarely governed by a single law: the GDPR, the Data Act, the Data Governance Act, the European Health Data Space and the AI Act impose obligations on overlapping sets of actors, and no established mechanism records, at the point of disclosure, which of them governs a given transaction, how their demands combine, and what the receiving party inherits. We work this problem through one running example, an individual's glucose readings moving from a general practitioner's care into the permitted secondary use of training and evaluating a clinical model, and derive from it six explicit legal requirements a machine-readable record must satisfy. We check these requirements against a model and set out the decentralised architecture that lets a vault produce the resulting record on its own. Most requirements are already met by writing the model's existing policy vocabulary specifically enough. But a record cannot show which access right an individual exercised. And because the Health Data Space groups training, testing and evaluation under one permitted purpose, it cannot distinguish what was requested from what was granted, nor which processing stage a release concerns. Recording these facts in a signed, purpose-scoped envelope, while trusting the determinations only a Health Data Access Body has standing to make, narrows the gap between being technically authorised to receive data and being demonstrably compliant in using it. Finishing the task across further instruments, jurisdictions and data types is the work this paper leaves to the agenda it sets out.

Keywords: trusted data exchange · cross-regulatory compliance · data governance

1 Introduction

No one mandates a wallet because they love wallets. A wallet is designed because someone has to carry the things that matter, and because who does the carrying

decides what is kept, what is shown, and to whom. That, at bottom, is what the European Union has been pursuing since its 2020 strategy for a single market for data [15], pursued at first mainly through the Data Governance Act and the Data Act, and updated in 2025 by a Data Union Strategy that turns the same ambition toward making high quality data available for artificial intelligence while reducing the cost of complying with the rules already in force [16]. Both strategies aim to protect data but to let it generate value, for the individual it concerns and for whoever puts it to further use. Whether the instruments adopted so far actually achieve this, or whether they remain better suited to protecting data than to letting anyone draw value from it, is a question this paper returns to throughout, and one a properly built infrastructure has a genuine chance of answering well. A feat that requires both technical and social input en masse.

Two more recent instruments make the ambition concrete for the case we develop. The EU Digital Identity Framework requires every Member State to place a certified wallet in its citizens' hands by 24 December 2026 [18]. The European Health Data Space gives that same citizen a central role in decisions over their own health data, both for the care they receive and for the considerably wider range of secondary purposes, among them research, but also policy making, regulatory activity and product innovation, that data may support once shared under a permit [5]. Plainly, these are a wallet regulation and a health data regulation. For what they enable, they both give the individual a genuine role in the governance of data concerning them, a role exercised alongside institutions and decentralisation, in whichever technical form it takes, is one means through which that role can be exercised in practice.

This paper takes as its instantiation of that role the personal data vault, a space controlled by the individual in which they aggregate personal data from multiple sources and disclose it selectively, on terms they set [19,34]. Concretely, we build on Solid, a set of open Web standards designed for exactly this purpose [33]. Solid gives each individual a pod, an addressable personal data store, reachable like any other resource on the Web, whose access rules the individual alone sets, and to which any authorised clinician, application or researcher connects. A vault, in our usage, is this pod, or anything architecturally equivalent to it.

We situate the vault within *data space*, a term used narrowly in some accounts, as infrastructure enabling data transactions between named participants under a shared governance framework, and broadly in others, as an umbrella spanning data models, contracts and the institutional practice around them [8]. We adopt the narrower, governance-centred sense throughout: a decentralised ecosystem in which participants exchange data under jointly recognised conditions of sovereignty, trust and interoperability. A vault is, in this space, the tool through which whoever holds the role of data holder, data subject, or both, takes an active part in a space that also involves data users, intermediaries, authorities and institutions with their own standing. The conditions governing what a vault discloses are a specific case of the broader negotiation problem such spaces exist to solve.

One further commitment is worth making explicit here, because it shapes everything that follows: *exchange* goes beyond *transfer*. Transfer invites the image of a parcel changing hands for instance: a file copied from one machine to another, and centralised systems are typically built that way. A vault need not be. It can answer a query, satisfy a stated purpose, or feed a computation without ever relinquishing its own copy of the underlying data; what crosses the boundary is a derivation, a result, or a scoped and time-boxed permission to use the data where it sits [7,19]. Sovereignty here is the objective, and presents as the difference between a disclosure the individual can retract and one they cannot. Whatever mechanism is meant to carry legal obligations across a hand-off, then, must travel as readily with an access grant or a computed result as with a copied file. This is a requirement, as we show, that existing policy models do not yet meet.

The regulatory landscape a vault must operate within is not straightforward [7,19]. The GDPR [1], the Data Governance Act (DGA) [2], the Data Act [3], the EUDI Framework [18] and the EHDS [5] each impose obligations on overlapping sets of actors, expressed in legal prose, adjudicated by different authorities. Sure, overlaps can be noted where the Data Act defers explicitly to the GDPR where the two conflict, and the EHDS cross references the Data Governance Act for the international transfer of certain highly sensitive data. They, however, do not state for transactions touching more than one of them at once, which obligation governs, how a conflict between them is to be resolved, and what a receiving party inherits as a result, and no mechanism, machine readable or otherwise, currently fills that role [23].

Section 2 grounds this claim in a single running example where an individual's glucose readings, moving from a general practitioner's direct care of the patient to a permitted secondary use in health research and from there into the training and evaluation of a clinical model, to which we return throughout the paper.

For now, it is enough to note what authorisation alone cannot tell us.[3] A party can be correctly authorised at the technical level while remaining non-compliant at the regulatory one, because authorisation answers only *is this party allowed to see this*, never *under which of several simultaneously applicable obligations, and with what residual duties attached.*

We therefore ask: *under a decentralised model in which the individual retains effective control over their own data, how can a single exchange carry a machine readable record of which obligations govern it.* Answering this in turn means: *resolving conflicts among simultaneously applicable obligations, establishing what a receiving party inherits as a result, and doing both in a way that keeps compliance verifiable across every subsequent hand-off.*

We address this in three parts. First, we conduct a cross regulatory analysis of the motivating example, tracing precisely which obligations attach at each hand off, where they conflict, and where they fall silent. Second, we turn to standards

[3] We use authorisation and access control interchangeably throughout, meaning the technical decision of whether a given party may access a given resource, as distinct from the separate question of whether that access is lawful under whichever instruments apply.

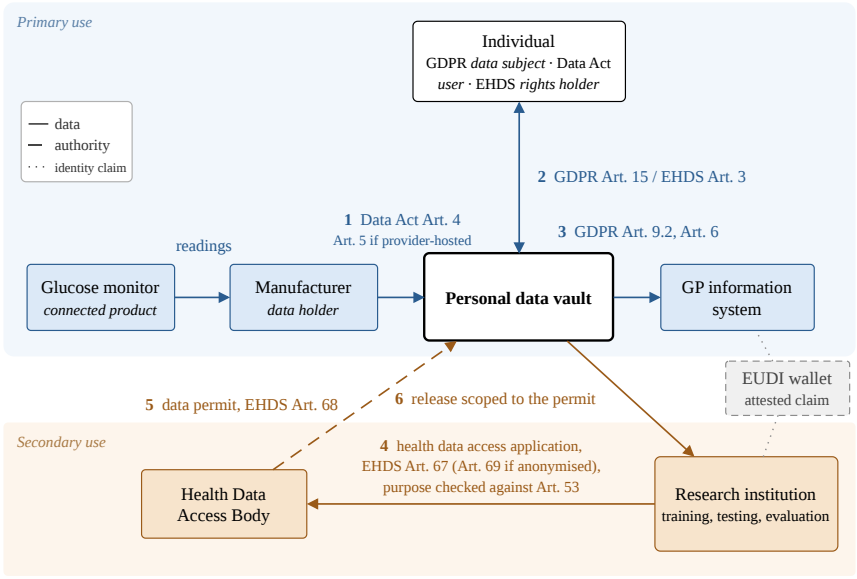

Fig. 1. Actors and flow of data in the motivating example, numbered in reading order. The individual and the vault are unshaded because they stand in both phases Step 1 is Data Act Art. 4 where the individual takes delivery and runs the vault themselves, Art. 5 where a provider hosts it, and only in the latter case does DGA Art. 12(a) attach.

already available for expressing permissions, purposes and obligations. Lastly, we set out how a decentralised architecture, of the kind already used in practice to separate authorisation from storage, needs to evolve to support this extended model, identifying the components such an architecture requires.

The remainder of the paper proceeds as follows. Section 2 develops the running example in full. Section 3 reviews the machine-readable policy models, decentralised architectures and regulatory instruments this work builds on, and states the three structural gaps that motivate it. Section 4 performs the cross-regulatory analysis proper. Section 5 presents the extended Trust Envelope model and the architecture built around it. Section 6 discusses limitations, revisits the economic and societal incentive gap in light of what precedes it, and situates the result within a broader techno-legal research agenda. Section 7 concludes.

2 Motivating Example

A single example runs through the rest of this paper, developed here in full and summarised in Fig. 1.

Vault ingestion. Consider an individual living with diabetes whose personal data vault aggregates data from two sources. The first is a continuous glucose monitor,

a connected product within the meaning of Data Act Rec. 14 [3], whose readings are recorded by the device and its associated service. The second is a copy of the clinical notes that form part of the health record their general practitioner is legally required to keep after each consultation.

The Data Act separates two routes into the vault. Under the Data Act Art. 4.1, when the data cannot be directly accessed by the user from their connected product, it is required that the data holder, here the manufacturer, makes readily available data available to the user. The individual takes delivery themselves and places the readings in a vault they run. In a scenario where the user would rather not do it themselves, Data Act Art. 5 is exercised. The user directing the data holder to make the same data available to a third party of their choosing, and a commercially hosted vault is such a third party. The distinction decides what attaches downstream. A commercially hosted vault may itself fall within the Data Governance Act's data intermediation services regime, binding the operator to the neutrality and structural separation duties of Art. 12(a), or, where it is run without reward for objectives of general interest, within the data altruism regime instead [2].

The clinical notes arrive by a third route. The individual obtains a copy by exercising the general right of access to personal data under GDPR Art. 15 and the more specific right of access to primary electronic health data the EHDS confers under Art. 3 [5], and holds that copy alongside the readings. Both rights precede any use the GP later makes of the data.

Because the readings are personal data, any processing of them requires a legal basis under GDPR Art. 6 from the moment they are recorded, and because health data is a special category of personal data, Art. 9 applies in addition, requiring one of its own exceptions before processing may take place at all [1].

Primary use. The GP subsequently draws on the vault-held readings and the health record to inform a consultation. This is processing of health data for the provision of care, resting on an GDPR Art. 9.2 exception together with an GDPR Art. 6 basis, and bounded by purpose limitation, GDPR Art. 5.1(b), and data minimisation, GDPR Art. 5.1(c): the GP may consult strictly what is needed. *User* under the Data Act, *data subject* under the GDPR, and *the person exercising* the EHDS *right of access* coincide here in one natural person, while the GP's system is a distinct data holder for the notes it generates. Role terms that look interchangeable across instruments are not, and their coincidence here is a property of this example.

Secondary use. Months later, a research institution seeks a pool of such records to develop a model that flags early signs of glycaemic instability. Under the EHDS this is *secondary use*, and the route depends on the form of the data sought. If the institution needs data in a form that identifies or may re-identify individuals, it submits a health data access application to the national Health Data Access Body under EHDS Art. 67. If anonymised statistical data would suffice, the institution submits a health data request under EHDS Art. 69 instead. In either case the purpose stated must be one EHDS Art. 53 recognises as a legitimate

ground for secondary use, here the development of a health-related algorithm. If satisfied, the Body issues a data permit under EHDS Art. 68, naming the purpose granted, the data covered and the period of validity. The permit is where the Health Data Access Body authorises a data user to receive data whose onward use it does not itself continue to police.

In this transaction, identity comes first. Every actor above must be identifiable to the others before any of it can happen, and from 24 December 2026 the EUDI wallet is the Union's answer, carrying the GP's clinical role and the institution's research accreditation as attested claims [18]. Purpose across stages follows: EHDS Art. 53 names training, testing and evaluation as a single recognised purpose, leaving open whether a permit granted for one development exercise still covers a later evaluation against a more recently collected cohort once the original processing has concluded. FInally, we assume throughout that the Data Act's user, the GDPR's data subject and the EHDS's rights holder are one person, where a parent managing a child's glucose data would separate them and Data Act Art. 5.7 would then require an independent GDPR Art. 6 basis before any sharing at all. Deployment lies past the example, and once the resulting system meets the AI Act's own definition its provider and deployer inherit obligations under that Act, scaled by risk [4]. We do not resolve that classification here.

3 Background & Prior Work

Three bodies of work converge on the problem this paper addresses. First, we provide an overview of relevant machine-readable vocabularies for EU regulatory obligations, and decentralised architectures that can deploy such resources for legal compliance purposes. Moreover, we describe the regulatory tools the motivating example already engages, before naming the gaps that persist.

3.1 Machine-Readable Vocabularies for EU Regulatory Obligations

Turning legal obligations into a resource a machine can interpret is still being actively explored and developed. The Open Digital Rights Language (ODRL), a W3C standard, supplies the language for permissions, prohibitions and duties, while being the one already being deployed in data spaces for access and usage control [24]. The Data Privacy Vocabulary (DPV) supplies taxonomies for purposes, legal bases, and processing categories, and is the de facto standard for representing data and AI-related legal requirements [29]. Several instruments now exist that could, in principle, carry a compliance record across a hand-off. Among them, the Trust Envelope model binds a unit of data to its provenance and its intended use as it moves, contextualising that binding through a history of where the data has been and a destiny of what it may still become, itself built on ODRL and DPV together [30]; alongside DUO-based approaches for health data sharing [28] and consent-focused ODRL profiles [12]. Regulatory instruments generate permissions, duties and prohibitions across a chain of actors; ODRL and DPV give that chain a machine-readable form; Trust Envelopes bind

the data to its provenance and use at each step. Recent work applies this same combination to specific instruments: to the Data Governance Act's data altruism and intermediation channels, modelling chains of access control policies that propagate downstream without any party expanding the permissions it received [27], to the AI Act, expressing an AI system's intended and precluded uses as permissions, prohibitions and duties [22], and to the EHDS, deriving a numbered set of legal requirements for health data exchange directly from the Regulation and modelling them the same way [11,13].

3.2 Decentralised Architectures

Personal data stores, of which a Solid pod is one instance, give the individual an addressable store whose access rules they alone set [19], built in Solid's own case on open Web standards [33]. Whether decentralising storage in this way actually returns control to the individual, or relocates the same information asymmetries somewhere less visible, remains a live and unresolved question in the literature [25,7], one that extends to the practical responsibilities such an architecture places on whoever governs a Solid pod, studied both for the GDPR specifically [20] and, more recently, for the multiparty access requirements the Data Act imposes [10]. Separately, recent work on modular usage control enforcement for data spaces names abstract components, independent of any single technology, that a decentralised architecture needs in order to authorise, decide, and continuously track access over time [6,31], a pattern Section 5 adapts directly. The EUDI wallet is a purpose built instance of the same underlying idea, applied specifically to identity credentials, and its own reference architecture defines the components and interactions such a wallet requires [14].

3.3 Regulatory Background

We focus on EU legislative pieces that concern data. The GDPR [1] governs the lawful processing of personal data generally, and applies throughout the motivating example from the moment health data is first recorded. The Data Act [3] governs access to and sharing of data generated by connected products, and defers to the GDPR wherever the two conflict, Art. 1.5. The Data Governance Act [2] governs neutral data intermediation and, as an alternative, data altruism for objectives of general interest. The EUDI Framework [18] governs identity attestation through a certified wallet. The EHDS [5] governs both primary and secondary use of health data specifically, and is designed to work alongside the other four, providing tailored rules for health data on top of the general data protection framework the GDPR establishes, a relationship prior analysis of the regulation sets out explicitly [11]. The AI Act [4] governs the development and deployment of AI systems, and its data governance duties attach once a system meets the Act's own definition, presupposing that a lawful basis for the training data already exists. There is a narrow connection between the DGA and the EHDS: the international transfer of highly sensitive non-personal health data

under EHDS Art. 88 is subject to conditions to be set out in a future Commission implementing act adopted under DGA Art. 5.13.

3.4 Gaps

Compliance tooling already exists for individual instruments taken one at a time, both in academic prototypes and in industrial deployments. We highlight a gap here: no existing model records which specific legal instrument source a given purpose or basis when multiple overlap during a single hand-off. Consequently, because these instruments simultaneously govern a personal vault, an operator holding several regulatory roles will struggle to determine which precise obligation applies to a given transaction. The same pattern holds for provenance. Solutions for recording provenance already exist and are well established, PROV-O [21] among them. However, what is missing is an account of how such provenance extends to hold a legal basis, a declared purpose and a governing permit together, across a hand-off spanning more than one regulatory regime at once. Prior prototyping of purpose-bound access control for medical data sharing shows the shape of this gap concretely [32], and current scholarship on the overlap between the AI Act's own data governance duties and the GDPR is mapping a related, third gap [23], namely how these instruments are meant to be read together at all once one presupposes the other. Lastly, machine-readable policies reduce the compliance burden for an actor already trying to comply, but say nothing about whether decentralisation changes who captures the value the data goes on to create [25,26].

4 Cross-regulatory Analysis

This section draws from the motivating example the specific legal requirements a trust envelope must satisfy. It builds from the simplest hand-off toward the most demanding one, and shows where those requirements draw on more than one instrument at once. Building from primary use toward secondary use, six requirements emerge.

R1. An identifiable GDPR Art. 9.2 exception, together with a GDPR Art. 6 basis and the purpose and minimisation limits of Art. 5.1(b) and 5.1(c), must be checkable independently of anything that happens to the data afterward.

R2. The record must show which of GDPR Art. 15's general right of access or EHDS Art. 3's specific right of access, or both, was actually exercised.

R3. Whoever discloses data must be named together with the basis relied upon, and the Data Act's two routes must be distinguished: a data holder making data available to the user under Art. 4.1, or a user directing that data to a third party under Art. 5, with DGA Art. 12(a) attaching to a provider-hosted vault in the second case alone.

R4. The request that opens a secondary use must be recorded as such, an access application under EHDS Art. 67 or a data request under Art. 69, together with the purpose it states, which must be checked against the purposes EHDS Art. 53 recognises before any permit issues.

R5. The permit issued under EHDS Art. 68 must be recorded with the purpose actually granted, the data covered and the period of validity, and, because Art. 53 groups training, testing and evaluation under one purpose, with the specific processing stage a given release concerns held separately from both the requested and the granted purpose.

R6. Duties attached to a grant, and the independent GDPR Art. 6 basis Data Act Art. 5.7 requires whenever the Data Act right holder is not the data subject, must remain checkable for as long as the grant is valid.

Primary use generates the first two requirements. The GP's consultation rests on GDPR Art. 9.2, which requires an identifiable exception before health data may be processed at all, together with GDPR Art. 6, and this basis must be checkable independently of anything that happens to the data afterward (R1). The individual's own access to the same data is a distinct legal event from the GP's use of it: GDPR Art. 15 grants a general right of access to one's own personal data, and EHDS Art. 3 grants a more specific right of access to one's own primary health data, and a record that cannot show which of these two rights, or both, was actually exercised has not captured what happened (R2).

Vault ingestion generates R3. The monitor's manufacturer discloses either by making data available to the individual under Data Act Art. 4.1, who then places it in a vault they run, or by directing it to a provider-hosted vault at the individual's request under Art. 5, and only in the second case does DGA Art. 12(a) bear on the vault operator; the record must show which route was taken and name the party that took it (R3).

Secondary use generates R4 and R5. The institution's request (e.g. "research"), an access application under EHDS Art. 67 or a data request under Art. 69, states a purpose the Health Data Access Body checks against the list EHDS Art. 53 recognises (R4). The permit the Body then issues under Art. 68 names the purpose actually granted, the data covered and the period of validity (R5). Because EHDS Art. 53 groups training, testing and evaluation under a single purpose, and because a request is routinely broader than the grant answering it, a record that flattens the two cannot later show which stage a given release served. Therefore, the requested purpose, the granted purpose and the stage must be held as separate, comparable facts.

Finally, any duty attached to the grant, a duty to notify on further use, a duty to honour retraction, and the independent GDPR Art. 6 basis Data Act Art. 5.7 requires whenever the person exercising the Data Act right is not the data subject, must remain checkable for as long as the grant is valid (R6).

These requirements intersect in specific ways. R1 and R4 both turn on GDPR's own purpose concepts, but R1 is answered by the GDPR alone while R4 needs the GDPR and the EHDS read together, since an EHDS permit only operates alongside a GDPR basis. R4 and R5 meet at the exact point the EHDS Art. 53's

drafting creates the problem it does: naming training, testing and evaluation as one purpose is precisely what makes a flat purpose record insufficient once evaluation draws on data collected after the original processing period closed. R3 and R6 meet where the Data Governance Act and the Data Act touch: DGA Art. 12(a)'s neutrality duty and Data Act Art. 5.7's independent basis requirement both concern who must check what before data changes hands, at different points in the same chain.

Regarding R5 and R6, upon commencing training, the research institution is subject to specific limitations. The AI Act's data governance duties and GDPR purpose limitation both continue to bind it. What changes is how easily that compliance can be checked from outside. A GP's consultation is a single, bounded act, simple to check against a stated purpose at the moment it happens. Training and evaluation extend over time, draw on pooled data from more than one source, and are not directly observable by the Health Data Access Body once its own permit process has concluded. That is a gap in what can be verified.

A similar precision applies to the Health Data Access Body's permit and to DGA neutrality. Both perform substantive work before access is granted: the Body checks a stated purpose against EHDS Art. 53's list before issuing a permit under EHDS Art. 68 and DGA Art. 12(a) bars a data intermediation service from becoming a data user of the data it intermediates. We notice that neither mechanism continues monitoring what the data user actually does with the data over the life of the resulting grant. We also see an absence of a record, produced at the point of disclosure itself, that a later reviewer, the Body, an auditor, or the individual, could consult without reconstructing it after the fact from whatever documentation the institution happens to have kept.

Once training gives way to a system meeting the AI Act's own definition of an AI system, its data governance duties attach as well. These duties presuppose a lawful basis for the training data, and whether that basis was in fact valid can be addressed by R1, R4 and R5, a distinction current work on the overlap between the AI Act and the GDPR is mapping in detail [23].

5 Decentralised Architecture for Trust Envelopes

The analysis above named six requirements. This section checks each against the Trust Envelope model as specified and sets out the components a vault needs to produce and consume the result.

5.1 Extending the Trust Envelope Model

A Trust Envelope binds a unit of data to the terms under which it travels [30]. It names exactly one sender and may name a data subject or data holder beside them; it carries a policy, written in ODRL with DPV terms for purposes and legal bases [24,29,11]; it carries provenance, the history of where the data has been and the destiny of what it may still become; and it carries the signatures

that bind these together. Three of our six requirements are met by writing those existing parts specifically enough.

R1's legal basis is a DPV term on the policy. R3's disclosing party is the envelope's sender, already mandatory and already distinguishable from the data subject. R6's duties are a native ODRL rule type whose validity window is expressible with existing temporal vocabulary [30]; the role divergence Data Act Art. 5.7 addresses is detectable by comparing the sender already recorded against the data subject already recorded, with no new term.

Three requirements need additions. R2 asks which right of access was exercised. DPV names the concept through its rights-exercise taxonomy [29], but nothing attaches an instance of it to a disclosure; we propose a property on the envelope's provenance that does. R4 and R5 divide between them what ODRL already separates: a request and an agreement are different things, but an envelope carries only the agreement. The granted purpose R5 names is the policy's own DPV purpose and needs nothing new; the requested purpose R4 records does, so we propose a request property linking an envelope to the request that preceded the policy it carries, and a stage term on the granted permission for the processing stage R5 holds separately.

5.2 Architectural Design

In a decentralised setting, where the data sits and who decides whether it may leave are separate concerns, so a vault can answer a request without giving up control of the data behind it. We describe the components such an architecture needs, shown in Figure 2, by what each contributes to the envelope, taking identity verification as a prerequisite: a requester presents an attested claim, a clinical role for the GP, a research accreditation for the institution, before any component below acts.

Three components stand between a request and a decision. When a request arrives, it must be evaluated against the policies that govern it. A retrieval component fetches those policies, the permit, consent or agreement, and filters them to the ones bearing on this request. A context component then assembles the facts the decision turns on but the policy does not itself contain: whether a valid permit exists and remains in force, which processing stage the request concerns, and whether the requester coincides with the data subject or diverges from them in the way Data Act Art. 5.7 addresses. A decision component weighs the request, the retrieved policies and that context together. Its output is not confined to grant or refuse: where the requested purpose is broader than the permit allows, it returns a demand that the requester restate its purpose in the permit's own terms, and grants only once the two match, scoped to the data the permit covers.

Once a grant is reached, an assembly component builds the envelope and signs it [30]. It produces the links to: the granted policy, the request that came first, the activity at hand, the rights-exercise record where a data-subject right was exercised and the signatures that bind the whole. Beyond the policy, this component assembles the envelope's provenance: which data was released, from

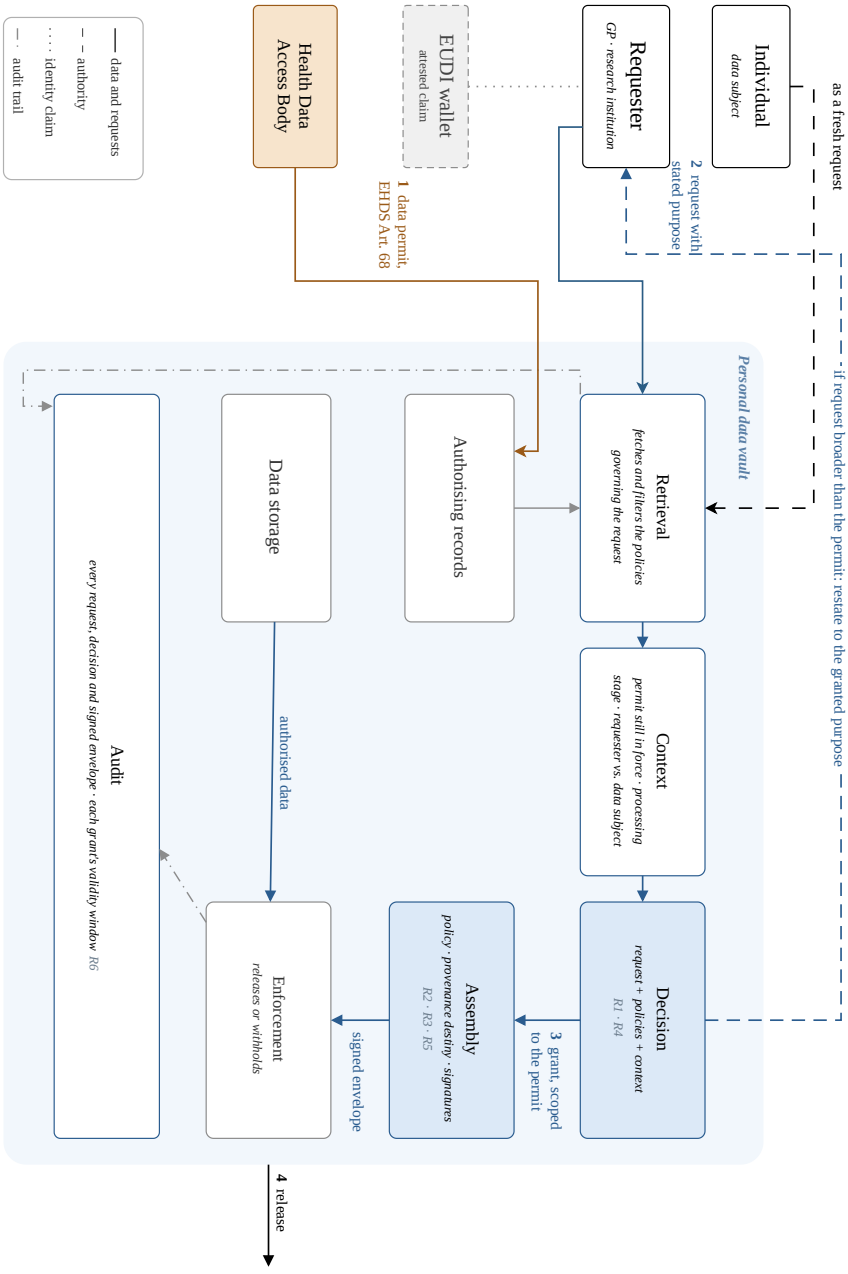

Fig. 2. The vault-side components, numbered in reading order, and the requirements each discharges

where, under whose authority, and, where a data-subject right was exercised, the record of that exercise. An enforcement component then releases the authorised data together with its signed envelope, or withholds it before data leaves the vault.

An audit component records every request, decision and signed envelope, and watches each grant's validity window. A retraction arrives directly from the data subject and re-enters the flow from the top: it is retrieved, contextualised and decided like any other request, and answered by a superseding envelope recorded in the same trail, so the history of a grant, including its withdrawal, is itself signed and auditable.

5.3 The Motivating Example, Materialised

Return to where the paper started. The institution's purpose was already narrowed once, at the Health Data Access Body, which issued a permit naming a specific purpose. When the institution then approaches the vault and states research, the decision component finds that broader than the permit it holds and demands the permit's own language. The institution restates, access is granted, scoped to the part of the record the permit covers, and the assembly component builds and signs the envelope. It carries the granted policy and the broader request that came first, so if the institution later processes beyond what was granted, the envelope is the evidence of what was actually allowed. A Turtle instantiation of this envelope, carrying the request, the granted policy and the rights-exercise record together, is available in the paper's companion repository.[4]

6 Discussion

GDPR Art. 30 requires a controller to keep records of its processing, and the EHDS requires a data user to report on what it did. However, an imbalance persists in the chain of custody. We do not observe an artefact produced at the point of disclosure itself that states what was granted, to whom, and for which stage, in a form a later reviewer could check without reconstructing it from whatever each party happened to keep. The EHDS provides an opportunity here. It already requires secure processing environments to keep auditable logs (Art. 73), and implementing acts specifying them are due by 26 March 2027 [5]. What we describe is a candidate shape for that specification: a record produced at the moment of disclosure, before processing begins. A data user must publish an anonymised results report within the period the Regulation allows, but that report describes what was done after the fact. Against the running example once again, it remains unclear what point would allow the individual living with diabetes to discover whether the model trained on their glucose readings was later tested for bias against patients like them, using their own data for a purpose the original permit never named. No party holds a record specific enough to verify this.

[4] `https://github.com/mudoki/trust-envelope-ehds`

Our architecture shows how the Health Data Access Body obtains a trace of its own permit's actual effect, produced automatically at the point of disclosure. This closes the gap that post-hoc reporting leaves open, because the trace exists before research begins. The research institution also holds a defensible record. For instance, if a regulator later asks whether training stayed within the permit's stated purpose, the institution can point to an envelope generated independently of its own records. Crucially, the individual gets the first document in this chain that records what was requested and what was granted as two distinct, comparable facts.

In a data space with different actors, systems and rules, costs of knowing what happened to a specific exchange once it is over are lowered, for the individual, for the body that authorised it, and for whoever received it. This speaks to the tension the introduction raised and left open, whether the EU's data strategies protect data or let it generate value. This knowledge allows for both protection and value creation. An architecture that makes it cheaper to know removes one of the reasons that outcome currently goes unrecorded at all.

6.1 Limitations

The running example lets the Data Act's user, the GDPR's data subject, and the EHDS's rights holder coincide in one person throughout. When a vault is maintained on behalf of a dependant, Art. 5.7 is invoked, requiring a valid GDPR Art. 6 basis before any sharing at all, and the negotiation would need a further claim establishing the requester's standing to act on the data subject's behalf, something this architecture is yet to develop.

The extended model rests on choices this paper does not fully discharge. Regarding R3 and R5, whether the entity that issues an envelope and the entity whose authority makes a disclosure lawful are one role or two is a modelling decision the regulations are yet to settle and we leave this open. A duty recorded in an envelope states what is owed. A policy evaluator can re-check it until it is discharged, provided the duty carries a deadline; where none is imposed it sits permanently unfulfilled without ever being violated, and specifying default deadlines is work we leave open.

We also leave open whether the resulting model counts as a high-risk system under the AI Act, a safety component of a medical device, or something else and the solution/response may change what an envelope needs to record at the training and evaluation stage in ways not anticipated in this narrative. Lastly, the Data Governance Act is the subject of a live Commission proposal to repeal it into a restructured Data Act [17]. A new Art. 4(a) of the AI Act, introduced through the AI Omnibus, permits providers and deployers to process special category data specifically to detect and correct bias, a purpose distinct from both training and evaluation and carrying its own safeguards [17]. A bias check run against a model like the motivating example would need recording as a distinct purpose that has been individually authorised.

6.2 The Economic and Societal Incentive Gap, Revisited

A purpose field and a duty to notify reduce the cost of demonstrating compliance for an actor already trying to comply. They do not settle whether decentralisation returns control to the individual or relocates the same asymmetry. Purpose negotiation protects the individual from an underspecified request. Done by hand it would raise the cost of participation for the actors least able to absorb it, a small research group or a hospital without a compliance function; machine-readable policies can be explored to keep that cost from landing on them. Nothing here changes the underlying economics of data as a non-rival good [26]. It offers infrastructure for expressing and enforcing an obligation once one exists. Whether any party comes to prefer creating that obligation in the first place remains unresolved, and one would need a bit more than infrastructure.

6.3 A Techno-Legal Research Agenda

The technical strand needs the architecture to support a client querying the vault before it knows which specific resource to target. We have assumed a resource specific request throughout this narrative arch, and the secondary use hand-off already gestures toward something vaguer, closer to a permit scale query. The legal strand needs the same cross-regulatory method applied here to one Health Data Access Body's permit extended to the Union's other health data access bodies. Twenty-seven national implementations are unlikely to converge unaided on what a stated purpose covers. A shared vocabulary is how that convergence can happen without waiting for the law to harmonise, and DPV's purpose taxonomy is an obvious candidate; extending the method here to further access bodies is the legal strand of the agenda [29]. The societal strand needs an actual account of whether an individual exercises a retraction right once it costs them nothing technically to do so. As this is a hypothetical scenario, we cannot demonstrate real deployment, and we point this as a limitation.

7 Conclusion & Future Work

This paper asked what it would take for a single data exchange to carry its own compliance record across a regulatory landscape no single instrument was built to span. The cross-regulatory analysis in Section 4 located where that record currently goes missing: at the point a Health Data Access Body's permit hands data to a research institution, and again at the point training gives way to evaluation within what the EHDS treats as one authorised purpose. Section 5 showed that three of the six requirements are met by the Trust Envelope model as it stands, and set out the three additions the rest require: a rights-exercise record attached to provenance and an explicit link from an envelope to the request that preceded it, and a term for the processing stage a release concerns.

What this already gives is a record produced at the moment of disclosure, holding what was requested, what was granted and which stage a release concerns

as separate facts, available to the Health Data Access Body, the data user and the individual alike. Future work extends it in three directions: a vault holder whose role diverges from the data subject's, deadline-bearing duties a policy evaluator can check to completion, and the cost of maintaining standing grants at deployment volume. The ground is also still moving, with the digital omnibus proposal and amendments elsewhere, and the purpose fragmentation this paper models is what those changes will test.

Acknowledgments

This project is funded by the European Union's Horizon Europe programme under the Marie Skłodowska-Curie Actions project HARNESS (grant agreement No.101169409).

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data (General Data Protection Regulation) (2016), https://eur-lex.europa.eu/eli/reg/2016/679/oj
2. Regulation (EU) 2022/868 of the European Parliament and of the Council on European data governance (Data Governance Act) (2022), https://eur-lex.europa.eu/eli/reg/2022/868/oj
3. Regulation (EU) 2023/2854 of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act) (2023), https://eur-lex.europa.eu/eli/reg/2023/2854/oj
4. Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) (2024), https://eur-lex.europa.eu/eli/reg/2024/1689/oj
5. Regulation (EU) 2025/327 of the European Parliament and of the Council on the European Health Data Space (2025), https://eur-lex.europa.eu/eli/reg/2025/327/oj
6. Akaichi, I., Slabbinck, W., Rojas, J.A., Van Gheluwe, C., Bozzi, G., Colpaert, P., Verborgh, R., Kirrane, S.: Interoperable and continuous usage control enforcement in dataspaces. In: Proceedings of the Second International Workshop on Semantics in Dataspaces (SDS 2024) (2024), https://research.wu.ac.at/ws/portalfiles/portal/69985704/SDS24.pdf
7. Asgarinia, H., Chomczyk Penedo, A., Esteves, B., Lewis, D.: "who should I trust with my data?" ethical and legal challenges for innovation in new decentralized data management technologies. Information **14**(7), 351 (2023). https://doi.org/10.3390/info14070351
8. Bacco, M., Kocian, A., Chessa, S., Crivello, A., Barsocchi, P.: What are data spaces? systematic survey and future outlook. Data in Brief **57**, 110969 (2024). https://doi.org/10.1016/j.dib.2024.110969
9. Belhajjame, K., et al.: Using a suite of ontologies for preserving workflow-centric research objects. vol. 32, pp. 16–42 (2015)

10. Coenen, T., Fierens, M., Esteves, B., De Loof, E.: A perspective on the use of personal data stores in data spaces to support multiparty data access as required by the data act. Data in Brief **62**, 111949 (2025). https://doi.org/10.1016/j.dib.2025.111949
11. Esteves, B., Dedecker, R., Slabbinck, W., Pattyn, F., Verborgh, R.: Semantic framework for legally-aligned health data exchanges. In: 2nd Workshop on ODRL and Beyond (OPAL 2025), co-located with ESWC 2025 (2025), https://ceur-ws.org/Vol-3977/OPAL2025-9.pdf
12. Esteves, B., Pandit, H.J., Rodríguez-Doncel, V.: ODRL profile for expressing consent through granular access control policies in Solid. In: 2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). pp. 298–306 (2021). https://doi.org/10.1109/EuroSPW54576.2021.00038
13. Esteves, B., Rodríguez-Doncel, V., Pandit, H.J., Lewis, D.: Semantics for implementing data reuse and altruism under EU's data governance act. In: Knowledge Graphs: Semantics, Machine Learning, and Languages, pp. 210–226. IOS Press (2023). https://doi.org/10.3233/SSW230015
14. European Commission: The European Digital Identity Wallet Architecture and Reference Framework. eIDAS Expert Group, https://digital-strategy.ec.europa.eu/en/library/european-digital-identity-wallet-architecture-and-reference-framework
15. European Commission: A European Strategy for Data. COM(2020) 66 final (2020), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0066
16. European Commission: Data Union Strategy. COM(2025) 835 final (2025), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2025:835:FIN
17. European Commission: Proposal for a regulation amending and repealing elements of the digital legislative framework (Digital Omnibus). COM(2025) 837 final (2025), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52025PC0837
18. European Parliament and Council: Regulation (EU) 2024/1183 establishing the European Digital Identity Framework (2024), https://eur-lex.europa.eu/eli/reg/2024/1183/oj
19. Fallatah, K.U., Barhamgi, M., Perera, C.: Personal data stores (pds): A review. Sensors **23**(3), 1477 (2023). https://doi.org/10.3390/s23031477
20. Fierens, M., Pandit, H.J., Tamo-Larrieux, A., Garcia, K.: A Solid use case to empower and protect data subjects: Responsibilities under GDPR for governance of personal data stores. Computer Law & Security Review **57**, 106133 (2025). https://doi.org/10.1016/j.clsr.2025.106133
21. Gil, Y., Miles, S.: PROV model primer (2010), https://www.w3.org/TR/prov-primer/
22. Golpayegani, D., Esteves, B., Pandit, H.J., Lewis, D.: AIUP: an ODRL profile for expressing AI use policies to support the EU AI act. In: 20th International Conference on Semantic Systems (SEMANTiCS 2024) (2024), https://ceur-ws.org/Vol-3759/paper17.pdf
23. Holtz, H.M., Ledendal, J.: AI data governance: Overlaps between the AI act and the GDPR. Law, Innovation and Technology pp. 380–409 (2026). https://doi.org/10.1080/17579961.2026.2633677
24. Iannella, R., Villata, S.: ODRL information model 2.2. W3C Recommendation (2018), https://www.w3.org/TR/odrl-model/
25. Janssen, H., Cobbe, J., Singh, J.: Personal information management systems: A user-centric privacy utopia? Internet Policy Review **9**(4) (2020). https://doi.org/10.14763/2020.4.1536

26. Jones, C.I., Tonetti, C.: Nonrivalry and the economics of data. American Economic Review **110**(9), 2819–2858 (2020). https://doi.org/10.1257/aer.20191330
27. Kisia, M., Kurteva, G., Verborgh, R., Esteves, B.: ODRL policy chains for the DGA's altruism and intermediation frameworks. In: OPAL 2026: 2nd Workshop of ODRL and Beyond, co-located with ESWC 2026 (2026)
28. Pandit, H.J., Esteves, B.: Enhancing data use ontology (DUO) for health-data sharing by extending it with ODRL and DPV. Semantic Web (2024). https://doi.org/10.3233/SW-243583
29. Pandit, H.J., Esteves, B., Krog, G.P., Ryan, P., Golpayegani, D., Flake, J.: Data privacy vocabulary (DPV) – version 2.0. In: The Semantic Web – ISWC 2024. pp. 171–193 (2024). https://doi.org/10.1007/978-3-031-77847-6_10
30. Slabbinck, W., Esteves, B., de Mildt, M., Dedecker, R., Rojas Meléndez, J., Verbrugge, S., Colle, D., Colpaert, P., Verborgh, R.: Incentivizing sustainable data exchanges through unique contextualization of history and destiny. In: Joint Proceedings of WOP-HAIBRIDGE 2025, co-located with ISWC 2025 (2025), https://ceur-ws.org/Vol-4093/paper6.pdf
31. Slabbinck, W., Termont, W., Dedecker, R., Esteves, B.: From access control to usage control with User-Managed Access (2026), https://arxiv.org/abs/2601.18761
32. Termont, W.: Authorization for data spaces. Tech. rep., Knowledge on Web Scale (KNoWS), IDLab, Ghent University – imec (2025), https://spec.knows.idlab.ugent.be/A4DS/L0/latest/
33. Verborgh, R.: Re-decentralizing the Web, for good this time. In: Seneviratne, O., Hendler, J. (eds.) Linking the World's Information: Essays on Tim Berners-Lee's Invention of the World Wide Web, pp. 215–230. ACM (Sep 2023). https://doi.org/10.1145/3591366.3591385, https://ruben.verborgh.org/articles/redecentralizing-the-web/
34. Verbrugge, S., Vannieuwenborg, F., Van der Wee, M., Colle, D., Taelman, R., Verborgh, R.: Towards a personal data vault society: an interplay between technological and business perspectives. In: Proceedings of the 60th FITCE Communication Days Congress for ICT Professionals. IEEE (Sep 2021). https://doi.org/10.1109/FITCE53297.2021.9588540