openEHR and HL7® FHIR® in Europe

Evidence, Architecture and Procurement Implications

Desiderata

Document 6 · Dr Kai U. Heitmann · Release 1.54.1 – 7 September 2026

What the communities would have to supply for the next version of this assessment to say more, or to say something different.

Written for: The standards organisations and their communities, the bodies that determine specifications, procurers and providers, vendors, and the researchers and funders who could close the evidence gaps the assessment had to record as open.

Documents 1 to 5 answer a procurement question with the evidence available in 2026. This document, by contrast, addresses those who could change that evidence.

Contents

Preamble

Where the evidence for these items sits. This document circulates on its own, and none of it is self-supporting. Every gap named here was established in the assessment set, and the set is published separately:

- **Document 4, the Evidence Annex,** is the primary reference. It carries the case material, the source criticism and the graded source ledger, in which every load-bearing source is recorded with the basis on which it was assessed, the proposition it supports, its evidence grade, the interests bearing on it and its retrieval status, followed by the full bibliography.
- **Document 3, the Assessment,** contains the derivation: the five architectural arrangements and the count of governed model crossings (§ 1.6.1), patient safety as a dimension of the mapping question (§ 1.6.2), the decision boundary for new infrastructure (§ 1.6.3), the distinction between conformance and interoperability (§ 1.5), the self-description of instance data and binding strength (§ 2.5), and the contested points weighed in § 2.6 and § 2.7.
- **Document 5, the Technical Report,** supplies the counted German profile corpus, what the packages do and do not carry, the terminology-binding distribution and the record of national profile governance, including the conformity-assessment procedure.
- **Documents 1 and 2** carry the decision and the executive argument, and they are the reason this list exists: what they cannot state, because the evidence does not carry it, is what the items below ask for.

Where an item states a figure — the 615 observations across 21 manually inspected C-CDA documents, the counted corpus or the alert-acceptance measurements — the source, its grade and its limits are in Document 4 rather than repeated here.

What this list seeks to achieve

The assessment set answers a procurement question with the evidence available in 2026. In doing so it had to record, repeatedly, that the evidence does not exist, that a mechanism is missing, or that a question cannot be settled from published material. Those points are not rhetorical concessions. Each of them is something a community could actually supply, and this list states what that would take.

A desideratum here is therefore not a wish in the general sense. It names who could act, what is missing, why it matters for patients, providers or researchers, and — the part that makes the list usable — **what would count as met**. An item without a test would be an opinion; an item with one can be checked at the next release, and can fail.

Three kinds of item, marked as such. The list mixes evidence gaps, missing artefacts and recommendations of good practice, and without a label a reader cannot tell which is which. Each item therefore keeps its own single test and carries one of three marks above it.

KIND	WHAT IT MEANS	ITEMS
Evidence needed	Something is not known and could be found out. A study, a count or a measurement would close it.	3, 7, 8, 21, 22, 23, 24
Artefact or mechanism needed	Something does not exist and could be built or published — a specification change, a test suite, a library, a register.	1, 2, 4, 5, 6, 9, 11, 13, 14

KIND	WHAT IT MEANS	ITEMS
Recommended practice pending evidence	The evidence to settle the question is absent, and this is what a careful party would do in the meantime. These are not research gaps and are not claimed to be.	10, 12, 15, 16, 17, 18, 19, 20

The third group is the one to read with the most caution. Its items rest on reasoning and on cases rather than on measurement, and if the evidence the first two groups ask for ever arrives, some of them will need revising or dropping. They are included because a procurement decision has to be made before the evidence exists, not because the evidence exists.

The list takes no side between the two standards. Several items are addressed to both communities in the same words, because the gap is the same on both sides. Where an item bears on one side only, that is stated as a fact about the artefacts, not as a criticism of the people who maintain them.

A. For the standards organisations and their communities

1. Conclude the binding-strength gradation in openEHR

Artefact or mechanism needed

FHIR distinguishes `required`, `extensible`, `preferred` and `example`, and a validator can act on the distinction. ADL and the Archetype Object Model carry an optional `term_bindings` section with no strength attached, so an archetype cannot express whether a consumer must honour a binding. The gradation was proposed inside openEHR in August 2020, with the FHIR model named as the template, and the discussion runs into 2023 without a visible conclusion. Until it lands, a safety-critical binding in an operational template has to be enforced by contract rather than by the model. *Met when* a released AOM/ADL version carries a strength on term bindings and at least one tool validates against it.

2. Publish mapping specifications and their test corpora, not only mapping tools

Artefact or mechanism needed

Tooling for openEHR-to-FHIR mapping now exists as a named, reusable specification with an executing engine, and the same is true on the FHIR-to-OMOP and OMOP-to-FHIR routes. What the assessment did not identify is a publicly available body of material that would let a third party judge how much meaning survives: representative clinical test sets, the expected result for each case, and the record of what was dropped.

The OECD report of 2026 describes mapping as a recurring task along the whole health-data lifecycle and calls for compatible mappings between standards. It supplies neither test corpora nor a measure of preserved meaning. It therefore confirms that this desideratum matters without meeting it (Document 4, E-10). *Met when* at least one mapping specification is published together with a clinical test corpus, an expected-output set and a documented list of the concepts it cannot carry.

3. Measure semantic loss instead of asserting it

Evidence needed

The assessment identified one quantitative examination of C-CDA output: 615 observations across 21 manually inspected documents from 21 technologies, with erroneous or missing terminology the second-largest category. The examination measures document heterogeneity, not patient harm, and not the loss of any particular openEHR-to-FHIR mapping. For openEHR to FHIR, and for both directions between FHIR and OMOP, no comparable quantitative examination was found in the material reviewed. The gap is therefore search-bound, but concrete.

How readily an unmeasured assumption of loss becomes an apparent finding is shown by the OECD itself: it presents mapping as a loss of quality and granularity without measuring any particular transition. What is defensible at that level is the risk. Whether relevant meaning is actually lost has to be established for a given source, target, use case and mapping (Document 4, E-10). *Met when* a preregistered study evaluates a defined source model, target profile and clinical use case against representative test cases, reports forward loss and round-trip loss separately, assesses clinical equivalence by qualified review rather than by field count, and classifies every material failure. A single proportion of concepts surviving is not enough: it weighs a laboratory unit and an allergy the same, and it hides which direction the loss occurred in.

4. Give the convergence process a deliverable that can be inspected

Artefact or mechanism needed

The joint annual meetings of the two organisations have produced a joint statement and an implementation guide in which an archetype accompanies a FHIR logical model as an informative document. That evidences convergence in publication. It does not evidence that transformation costs fall, which is the procurement-relevant claim. *Met when* a jointly maintained artefact — a mapping set, a shared value-set layer or a joint conformance test — has been used in at least two independent implementations, and the mappings, validation steps or maintenance tasks it removed are documented against a stated baseline.

5. Publish snapshots, not only differentials

Artefact or mechanism needed

Of the German artefacts held in published package form for the count in Document 5, only a small minority ship a resolved snapshot, so it cannot be established from the package whether an element left unbound in a differential is bound by its base definition. That single omission blocks an otherwise mechanical audit of terminology rigour.

Two boundaries belong on this request. Where only a source repository was available, the question was not put at all, a repository not being where a snapshot is expected, so this is a request about what packages carry and not a measurement of every publisher. And the corpus holds one package line that ships no snapshot at five successive versions and a full set at the sixth, which establishes that this is a publication setting rather than a limit of the tooling, and therefore something a publisher can simply switch on. *Met when* the major profile families publish resolved snapshots alongside differentials as a matter of course.

6. Treat tooling openness as part of the openness claim

Artefact or mechanism needed

A specification is open when its text is public. The practical openness of an ecosystem additionally depends on whether a competent team can model, validate, query and migrate with interchangeable tools. The two are regularly conflated, and the gap sits on both sides in different places.

For openEHR it is the sharper problem, because the most plausible way to lower the entry hurdle is for a vendor to encapsulate the modelling complexity in tools, and precisely that would make the standard open in principle while a single vendor owns the only practicable implementation path. For FHIR the tooling is broad and largely open where conformance is mechanical — profile authoring, validation, terminology operations — and thin where it is clinical: no open tool establishes that content produced against a profile is usable for a stated purpose at the receiving end.

The openEHR side of that gap is no longer an inference: of eleven production-capable openEHR repositories surveyed, **two are free to use and open source** (Delussu et al. 2024; Document 4, A-2). *Met when* — **for tooling openness** — a reference stack is published with which a project can author, validate, query and migrate clinical models end to end, and each step of it is available either as open source or from more than one supplier, with the paper stating which of the two applies at each step, because they are different guarantees and only the first

survives a supplier leaving the market. **Clinical end-to-end validation is a separate requirement and is at item 13**, where it belongs; it was previously attached here and made a single test out of two unrelated goals.

7. Measure reuse of the shared model layer, and publish the measurement

Evidence needed

Both communities describe their content models as reusable; in the material reviewed no ecosystem count is published for either side. For FHIR the measurement is mechanical: `baseDefinition` makes derivation a graph, and running it once over the German corpus shows reuse at roughly 48% of resource profiles, nearly two thirds of it inside a single package, with only 62 of 2,317 profiles naming a national base profile as parent (Document 5, D-10.7).

For openEHR the equivalent is equally computable, because a template names the archetypes it composes and an archetype names its parent, so the share of content drawn from the international library rather than modelled locally could be reported for any holding. No such count was identified in the material reviewed for this assessment.

Spain now plans the object this measurement needs. The ministry-commissioned Delphi recommendation asks for a central, open repository of clinical information models with federated editorial governance (Document 4, E-12). If it is built, the reuse share this item asks for becomes computable for a whole national system; the item itself is unchanged, because a repository is the precondition of the measurement, not the measurement.

Reuse is not a virtue in itself, and the item is not a call for more of it: every derivation couples the release cycle of the child to that of the parent, which is the same coupling the German governance record names as its central problem. What the measurement is for is that a community should know the shape of its own ecosystem before it argues about it. *Met when* at least one national or vendor holding on each side publishes a derivation or composition count over its own artefacts, with the method stated in enough detail to be repeated. Repeating it at a later release would additionally show a trend, and is worth doing, but a single transparent cross-section already closes this item.

8. Report how much of the source data the standard actually carried

Evidence needed

A systematic review of 99 studies applying FHIR, OMOP-CDM or openEHR between 2021 and 2024 found that only 27 of them state their coverage — how many variables of the source dataset were mapped into the standard — and that the figure is so rarely given for FHIR and openEHR that the standards cannot be compared on it (Marfoglia et al. 2026; Document 4, E-9). Coverage is the lowest-burden transparency measure available for a transformation: it needs no outcome study, no cost model and no comparison group, only a numerator and a denominator that the implementing team already knows.

Its absence is what allows a mapping to be described as working while the share of clinical content it drops stays unexamined. *Met when* implementation reports and project publications state the number of source variables, the number mapped, and what happened to the remainder. A funder or journal asking for it as a matter of course would make that routine, and is the mechanism rather than the evidence.

9. Decide whether FHIR is to have a curated library of reusable form modules, and say so either way

Artefact or mechanism needed

SDC specifies the composition mechanism — a `subQuestionnaire` extension referencing another Questionnaire by canonical URL, expanded by `$assemble` — but establishes no shared holding of modules, and notes only that research organisations keep question libraries of their own. That a curated holding is achievable is not speculative: the REDCap Shared Library reviews instruments for relevance, correct functioning and copyright before release and

lets projects import them. openEHR has such a library at the model layer; FHIR has the mechanism and no library at the form layer. *Resolved when* the community records an explicit governance decision either way, including a decision that form modules are left to jurisdictions and programmes. **Met when** a body publishes a curated set of Questionnaire modules with a stated review procedure and a resolvable canonical namespace. The two are deliberately separate: a decision not to build a library settles the governance question and leaves the reuse gap exactly where it was.

10. Reuse deliberately rather than maximally, and state the condition

Recommended practice pending evidence

The counted German corpus shows reuse concentrated inside programmes: 721 of 1,110 in-corpus derivations stay within one package, and ten hub artefacts carry 504 of them (Document 5, D-10.7). The tempting conclusion is that there should be more derivation across programme boundaries. It does not follow. A derivation couples the release cycle of the child to the parent's, and divergent release cycles are what the German governance record names as its central problem, so a deeper cross-family hierarchy would sharpen the difficulty it is meant to relieve.

What is defensible is a condition rather than a quantity: derive from a parent maintained on the same cadence or a slower one than the artefact deriving from it, and cross a package boundary only where a named body answers for the parent's lifecycle and for its breaking changes. The same discipline applies on the openEHR side to templates drawing on archetypes governed elsewhere. Underneath both halves of the condition sits a version question. In the German corpus 6.8% of parent references and 9.8% of value-set references carry a version, while the package manifests above them pin an exact version in 89.3% of cases (Document 5, D-10.7). That division is by design — inside a guide the manifest fixes the parent — but it leaves the artefact silent about the release it was written against, and it offers nothing at all where an artefact is read outside a guide, in a registry or on a server holding several families. *Met when* a profile family or template holding publishes the rule under which it derives from artefacts outside its own governance, names the body answerable for each external parent it depends on, and binds those external references to a version rather than to a moving target.

11. Annotate model nodes semantically, or record why not

Artefact or mechanism needed

A node labelled "body height" is readable and local; the same node annotated with SNOMED CT `1153637007 |Body height (observable entity)|` carries an identity that outlives the model it was written in. **Both standards offer a route to machine-processable semantic identity, and the routes are not the same shape.**

In openEHR the optional `term_bindings` section maps the node identifiers themselves — at-codes and id-codes — to external concepts, so any node in an archetype can be given an external identity. In FHIR a `binding` attaches to elements of a coded type and constrains their *permitted values* rather than annotating the node; annotating an element definition itself is a different mechanism, `ElementDefinition.code`, and in forms there is the further choice between an enumerated `answerOption` list and an `answerValueSet` bound to a published value set, with `item.code` available for the question (§ 2.5.2). The reach therefore differs, and it differs in openEHR's favour at the node layer.

Annotation discipline is therefore not a property of either standard but of whoever maintains the models, and it is exercised unevenly. The counted German corpus contains substantially more listed fixed and pattern constraints than explicit bindings; because those constraints were not classified by element path and datatype, the count does not establish how many of them govern coded clinical content.

A statutory body has asked for this as law. On 11 August 2026 the Interop Council asked that semantically standardised data holding at the logical layer be prescribed by statute, so that data are held with standardised semantic annotation, while expressly asking that no requirements be placed on databases or system architectures

(Document 4, E-11). The item below is unchanged by that: an annotation policy is what would make such a duty checkable, whoever imposes it.

Met when a model holding publishes an annotation policy that states which clinically meaningful nodes require an external concept identifier, how an exception is recorded, and how compliance is reviewed — and a reader can tell an annotated node, a deliberately unannotated one and an overlooked one apart without asking the authors. **An identity on every node carrying clinical content would not be attainable**: not every node has an atomic counterpart in a terminology, and a rule that cannot be satisfied is not a test.

B. For regulators and the bodies that determine specifications

12. Keep the exchange obligation and the persistence obligation apart, in the text

Recommended practice pending evidence

A duty to release data in a prescribed exchange format constrains the exposure layer of a system. It does not settle which model holds the record, and a procurement that reads it that way is drawing a conclusion the instrument does not supply — in either direction. Draft legislation whose summary and operative text diverge on this point invites exactly that reading.

The line has since been divided once more, and the item survives the division. The Interop Council asks that a statute reach the *logical semantic* layer of data holding while leaving databases and system architectures alone (Document 4, E-11). That is a third position, not a collapse of the distinction: it separates what is held from how it is stored. This item asks that whichever line is drawn be drawn in the operative text rather than in a summary, and a three-way line needs that discipline more than a two-way one, not less. *Met when* the operative provisions state the addressee, the layer and the format duty explicitly, and no summary text asserts a data-holding duty that no provision carries.

13. Extend conformity assessment beyond conformance

Artefact or mechanism needed

A national conformity-assessment regime with published test suites, issued certificates and a public positive list is the missing half of every governance argument: a specification without an assessment route binds no one in practice. What such a regime records, however, is that a system answered the cases in the suite. It does not record that the exchanged content was usable at the receiving end.

The OECD's cross-country comparison names uneven standards uptake, local extensions, vendor-specific endpoints and insufficient conformance testing among the recurring obstacles. That is an independent confirmation of the problem at policy level, but not evidence that any particular assessment regime achieves clinical interoperability (Document 4, E-10).

The body responsible for advising on the German assessment route concedes the same limit from the inside. Asking for terminologies, codes and value sets to be made binding across sectors, with LOINC and SNOMED CT as its examples, the Interop Council adds that a complete check within conformity assessment will not be possible in every case (Document 4, E-11). A regime whose own advisory body says it cannot verify everything it requires is the case this item is about. *Met when* an assessment regime adds at least one clinically defined end-to-end case and publishes the result per system, the case specifying all five of: the clinical use case, the sending and the receiving system, the content expected to arrive, the clinical acceptance criteria, and how a deviation is handled. Without those five a case can be declared passed on syntax alone, which is what this item exists to get past.

14. Resolve cross-family conflicts in the open, and keep a register of them

Artefact or mechanism needed

Where one binding profile family makes an element mandatory and another prohibits it on data-protection grounds, both conformant and both nominally on the same national base layer, the conflict is not a modelling accident. It is a governance product of separate determination powers, and it is invisible to anyone who reads only one family.

Two things make this actionable rather than merely regrettable. The conflict class is not national: an analysis of US Core 5.0.1 against the International Patient Summary 1.0.0 finds the same kind of collision between two of the most widely implemented guides in the world, one of the three resource areas examined infeasible outright (Kramer & Moesel 2023; Document 4, D-14).

And the detection is mechanical once profiles are matched by what they are meant to represent — that paper supplies the method, and the remaining manual step is reading profile narratives, which a computable expression of scope would remove. A body that wanted a register would not have to invent the instrument.

Avoiding contradictory requirements is now also asked for from the advisory side. The Interop Council asks that the competence centre and the BfArM make their terminology determinations within their respective remits and that contradictory requirements be avoided (Document 4, E-11). It names no register and no responsible body for collisions that arise anyway, which is what this item asks for.

Met when a public register lists known cross-family conflicts with their status, and a named body is responsible for closing them.

15. Require that value sets be resolvable, versioned and retrievable at the point of validation

Recommended practice pending evidence

A binding is worth what its resolution point is worth. A `required` binding against a value set that exists only inside an implementation is a declaration, not a constraint, and licence conditions that stop a validator resolving a code system at run time have the same effect as no binding at all.

The demand for binding terminologies now has statutory backing behind it, and it stops short of this item. The Interop Council asks that terminologies, codes and value sets be determined bindingly across sectors and applications, naming LOINC and SNOMED CT (Document 4, E-11). Determining a value set and making it resolvable at the point of validation are two different acts, and only the first is asked for there. *Met when* every binding in a nationally determined profile resolves to a published, versioned artefact that authorised implementers and validation services can retrieve under licence conditions that are stated openly and are workable in operation. **Free of charge is not the test**, because terminology licensing differs by country and by contract and a blanket demand for it would be unmeetable; the test is that the conditions are known in advance and do not stop a validator resolving a code at run time.

C. For those who procure and provide care

16. Ask the system-of-record question in the tender, and test it at acceptance

Recommended practice pending evidence

The decisive question for new infrastructure is not which store is bought but which structures hold the clinical content authoritatively, and whether everything else can be rebuilt from them. It is ordinary engineering, it is testable, and asking it separates products whose labels are indistinguishable. *Met when* tenders carry the

requirement, the evidence obligation and the acceptance test, and at least one award record shows the test being run rather than promised.

17. Inventory the safety-relevant configuration before a change of system, not after

Recommended practice pending evidence

After one examined change of system, the frequency and the acceptance pattern of drug–drug interaction alerts diverged sharply. The study measured no adverse drug events and isolated no single cause. What it shows more narrowly is that carrying clinical data across does not guarantee functional equivalence: alert tiering, hard stops, the moment at which a check fires and the acknowledgement logic can live in the application and in configuration. A migration plan that considers only the data mapping therefore draws the clinical risk too narrowly. *Met when* migration plans carry a configuration inventory, a specified target equivalent for each item, and a named list of functions that will not be reproduced — signed off before award.

18. Contract for the model repository, not only for the data

Recommended practice pending evidence

Whoever procures a model-based persistence layer procures an obligation: every model version needed to interpret the stored data must remain available for as long as the data are retained, or a verified migration must make the earlier model unnecessary. Retention alone is the default and may run to decades, depending on the statutory, contractual and clinical retention requirements that apply, which in Switzerland and Germany differ by data type, legal basis, care setting and canton. The same discipline applies to profiles, extensions and value sets on the FHIR side. *Met when* the contract names the artefacts, the retention period or the migration commitment that replaces it, the export format and the party responsible after the contract ends.

D. For vendors

19. State the architecture in the product documentation, in the terms the decision needs

Recommended practice pending evidence

Two unrelated suppliers market products described as FHIR-native whose own architecture descriptions place them at a different arrangement: content originates in primary systems, crosses one governed boundary at ingest and is then kept in FHIR. Nothing is wrong with that architecture. What is wrong is a market word that answers a question it does not answer. *Met when* product documentation states plainly where the clinical record is written, in which model it is kept authoritatively, and what is derived.

20. Make exit a documented feature rather than a negotiation

Recommended practice pending evidence

Vendor neutrality is a property of governance, intellectual-property rights, skills and exit provisions, not of the standard on the label. A published export path for data, models, mappings and terminology bindings, in documented formats, is the cheapest way for a supplier to defuse the lock-in objection. Where a specification already provides that path, implementing it is the shortest route: the openEHR EHR Extract exists for exactly this purpose and is declared by **one of eleven** surveyed openEHR repositories, with a second offering it through a separate product (Delussu et al. 2024; Document 4, A-2).

Where a standard is marketed on vendor neutrality and its own exit mechanism is implemented that rarely, the practical portability claim is insufficiently demonstrated at the product layer, whatever the specification provides.

Met when the export procedure is publicly documented and has passed an exit test covering data, models, mappings and terminology artefacts, the test being either independently witnessed or reproducible from the published procedure. **A named customer case is expressly not required**, because a supplier cannot describe one without the customer's consent and the test should not depend on that.

E. For research and its funders

21. Fund the comparison that is still missing

Evidence needed

The assessment identified no controlled head-to-head study comparing clinical outcomes or full lifecycle costs of otherwise comparable architectures, and reviews of FHIR in research do not include openEHR as a comparator. The absence is old enough that it is now quoted as if it were a finding about either standard, which it is not. A systematic review of all three standards published in 2026 closes by asking for exactly this — a rigorous assessment of process efficiency and economic impact — which makes the gap one that an independent team now names as well (Marfoglia et al. 2026; Document 4, E-9).

The OECD does not close this gap either. Its allocation of roles between FHIR, openEHR and OMOP is illustrative and not comparatively examined. Precisely because it now appears in an international policy framework, the missing direct comparison of architectures becomes more pressing rather than dispensable (Document 4, E-10).

The setting for the missing comparison now exists on paper. Spain's ministry-commissioned recommendation, if enacted, would run an openEHR record behind a FHIR exchange layer at national scale (Document 4, E-12), and the paper itself concedes its recommendations are not evidence-based. A programme of that size that funded the lifecycle-cost and outcome measurement alongside the build would close the gap this item names; the fulfilment test is unchanged.

Met when at least one registered comparative study reports lifecycle cost or a clinical endpoint for two architectures serving the same purpose in comparable settings.

22. Survey the FHIR repository product layer the way the openEHR one has been surveyed

Evidence needed

A peer-reviewed survey describes seventeen identified openEHR repository products against fifty-four questions; eleven vendors or developer groups responded, and the answer set is published (Delussu et al. 2024; Document 4, A-2). No methodologically comparable survey of the FHIR product layer was found in the material reviewed.

What exists is vendor documentation, a 2019 evaluation by a research network whose authors state they set the servers up from documentation alone, and a supplier benchmark in which the supplier's own product wins on every axis but one.

The available product-level evidence is therefore asymmetric, and asymmetric against expectation: the larger ecosystem has the less examined product layer, and a procurer comparing arrangements finds better published evidence about the option this assessment does not recommend by default. The openEHR survey is a usable starting point, though its questions would have to be adapted to FHIR servers, to repository capabilities and to a different product boundary. *Met when* an independent group publishes a comparative survey of FHIR repository products covering conformance to the specification, terminology handling, bulk export, version support and exit paths, with the instrument and the responses published so a vendor can contest an answer.

23. Publish the negative and the mundane

Evidence needed

Reviews in this field concede publication bias, quality assessment is often structural rather than substantive, and project reports that did not go well reach print rarely. The FIKS long-term case study is especially informative precisely because it recorded which assumptions failed. *Met when* funders require a published end-of-project report including the assumptions that did not hold, and the reports are indexed where implementers look.

24. Report cost in a form that can be compared

Evidence needed

Cost claims in this debate are either absent or incommensurable: a licence figure here, a project budget there, an FTE estimate somewhere else. A minimal common reporting frame — modelling, mapping, validation, upgrade, migration and supplier-change effort over a stated period, per use case — would make two projects comparable even where the absolute numbers are confidential.

The OECD report illustrates the problem. It places projected benefit figures from different countries and periods side by side and concludes that interoperability will finance itself in the long run. At the same time it reports, for Finland, an investment of up to two billion euro in harmonised data models based on openEHR and OMOP over three to five years. None of these figures separates the cost of the standards, the transformations and the organisational change, or compares them with an equivalent alternative architecture. As an order of magnitude they are relevant; as an architectural comparison they are not (Document 4, E-10). *Met when* two independent projects publish lifecycle cost against the same frame.

F. The outcome criterion across all evidence items

This is not a twenty-fifth desideratum. It is the measure against which several of the items above would have to prove themselves, and it is stated separately because no single one of them carries it. Everything above is instrumental: the reason to close these gaps is that clinical meaning should survive the journey between systems, that a safety function should not disappear because a hospital changed supplier, and that a person's record should remain interpretable long after the application that produced it has been retired.

None of those is yet measured consistently enough to compare the architectures examined here in terms of patient-relevant outcomes. Transformations and changes of system create additional validation and monitoring obligations; the literature reviewed quantifies neither the resulting patient harm nor a difference between one conversion and two. That is precisely why comparable endpoints are needed.

Met when at least one health system reports a patient-relevant indicator against an architectural change — medication-reconciliation discrepancies at admission, alert-acceptance behaviour after a change of system, or the completeness of a transferred record measured clinically rather than syntactically — and reports it whichever way the result falls.

A note on the standing of this list

These are desiderata, not findings, and none of them is a condition the assessment's recommendation depends on. The recommendation stands on the evidence available now; this list states what would have to appear for the next version to say more, or to say something different. Several of these items could strengthen, qualify or alter parts of the assessment if they were met — among them 1 and 5 on the artefact side, 3 on mapping loss, 21 and 24 on comparative outcomes and cost, 22 on product portability, and the patient-relevant endpoint of section F. That is why they are stated as testable desiderata and not as support for the present recommendation.

The assessment set

	DOCUMENT	WRITTEN FOR
1	**Decision Note**	boards, commercial and political leadership — the decision in one page
2	**Executive Brief**	CIOs, CMIOs, enterprise architecture, informed procurement
3	**Assessment**	interoperability and standards specialists, and anyone checking the derivation
4	**Evidence Annex**	reviewers checking cases, source criticism and the graded source ledger
5	**Technical Report: German FHIR Profile Governance**	the German FHIR profiling and terminology teams, and those responsible in governance structures
6	**Desiderata** (this document)	the standards communities, specification bodies, vendors, researchers and funders

Documents 1, 2, 5 and 6 are also available in German.

Method, and the use of a large language model. The literature search, the collation of sources, the drafting, and the consistency and readability checking across versions of this set were carried out with the assistance of a large language model. The editorial pass that set paragraph and sentence structure was made after the findings were fixed and did not change them. Not all of the cited literature was read in full: for part of it a summary extract was the working basis, and Appendix G of Document 4 records source by source which basis applies. Document 4 also states the limits of the method in full, including what a re-derivation of the evidence grades can and cannot establish. The grading rule, its application, the assessment and the conclusions are the author's, as is responsibility for any error that remains.

Transparency. The author is CEO of HL7 Deutschland, a Senior Expert and Community Event Manager at HL7 Europe, and a member of the ART-DECOR Expert Group. Those roles bear on a document that addresses requests to both standards communities, and they are stated for that reason. ***The list presented here is his personal professional view as an interoperability expert and does not represent an official position of HL7 Deutschland or HL7 Europe.***