Consistency Is Not Coherence: Orientation Search for Certified Alignments Between 4D Defence Upper Ontologies

Fabio Rovai[1]

[1]The Tesseract Academy (Kampakis and Co Ltd), 5 Brunswick Park Gardens, London N11 1EJ, United Kingdom

Abstract

We align three upper ontologies that sit under UK and NATO defence data infrastructure: the Information Exchange Standard (IES), the Higher Quality Data Model (HQDM) that underpins the National Digital Twin, and Basic Formal Ontology (BFO). No public alignment between IES and HQDM existed. Promoting a hand-curated 17-correspondence crosswalk to OWL and reasoning over the complete merged ontologies with HermiT produces three results that we believe matter beyond this pair. First, the published `hqdm.owl` shipped by GCHQ is *natively incoherent*: 39 of its 229 named classes are unsatisfiable before any alignment is added, while IES and BFO have none. An alignment evaluated against it inherits a broken target. Second, naive equivalence promotion leaves the merge *consistent*, so the usual check passes, while creating 100 new unsatisfiable classes and 218 entailed source-internal subsumptions that neither standard asserts. Consistency is the wrong acceptance test for an alignment. Third, we present *orientation search*, a repair operator that treats the direction of each correspondence ($\sqsubseteq$, $\sqsupseteq$, $\equiv$, discard) as the search variable rather than treating mapping deletion as the only move, with a reasoner as oracle for coherence and conservativity. It yields a 21-axiom IES↔HQDM bridge with zero new unsatisfiable classes and zero conservativity violations, in which every weakening carries a machine-found counterexample as its justification. Crossing to BFO is possible only through the occurrent branch: with the rest of the bridge fixed, mapping `ies:Entity` to `bfo:material entity` empties 101 IES classes, weakening that mapping to a subsumption still empties 100, and re-targeting it to `bfo:history` empties none. The defect is the target, not the strength. Finally, we test whether existing systems produce this bridge. LogMap and LogMapLt accept the crosswalk's documented false friend `ies:Event` $\equiv$ `hqdm:event`, map onto natively unsatisfiable classes, and carry 87 and 214 new unsatisfiable classes once read as logic; LogMap's repair, given our expert input, is safe but keeps none of the five provable equivalences; two LLM oracles score 6/12 and 7/12. Every system, classical and neural, falls for the same false friend the reasoner-refereed search avoids. All artifacts are open.

Keywords

ontology matching, alignment coherence, alignment repair, conservativity, upper ontologies, 4D ontology, defence data standards

1. Introduction

Alignments are usually published as sets of correspondences with a similarity score, and consumed as if they were logical axioms. The gap between those two readings is where this paper lives.

Our setting is deliberately unglamorous and operationally real. The UK Information Exchange Standard (IES) is the vocabulary in which UK defence and national-security systems exchange assertions about people, events and places. The Higher Quality Data Model (HQDM) [13] is the foundation data model beneath the National Digital Twin and the built-environment estate. Both descend from the same BORO and ISO 15926 4D tradition [14], so they ought to be cousins, and an autonomous system reasoning about a mission in a real place has to join what IES says about the operational picture to what HQDM says about the terrain. Basic Formal Ontology (BFO) [15, 16] is the third party in the room because it is the realist upper ontology of the US and NATO defence data foundry, so a UK-to-coalition join runs through it. No public machine-readable alignment between IES and HQDM

OM-2026: The 21st International Workshop on Ontology Matching, collocated with the 25th International Semantic Web Conference ISWC-2026, October 25th or 26th, 2026, Bari, Italy

fabio@thetesseractacademy.com (F. Rovai)

https://gov.tesseract.academy (F. Rovai)

0009-0001-7244-794X (F. Rovai)

existed before this work. For IES and BFO, a thirteen-author team spanning the IES, BORO and BFO communities has recently compared the two frameworks at the level of design patterns [19], explicitly deferring "axiomatic translation definitions" that would move "from plausible semantic equivalence to provable equivalence" to future work. This paper supplies that layer: every correspondence here is a reasoner-certified OWL axiom.

We built one, then asked the question that the ontology matching community has long treated as central and that alignment consumers still routinely skip: what happens when the correspondences are read as logic? The answer, on real published artifacts, is worse than the folklore suggests, and the repair is more interesting than deletion.

Contributions.

1. **A native-defect finding** (§4.1). The published `hqdm.owl` has 39 unsatisfiable classes with no alignment present, invisible both in the asserted hierarchy and to a consistency check. Any evaluation of an alignment *into* HQDM measures against a broken target, and any mapping onto one of the 39 is vacuously true and semantically empty.
2. **A separation result on real data** (§4.2). Naive equivalence promotion of a 17-correspondence expert crosswalk yields a merge that is consistent, and therefore passes the check most pipelines run, while adding 100 new unsatisfiable classes and 218 conservativity violations. We give the isolation: exactly one of the ten class targets is natively unsatisfiable, and weakening exactly that one edge takes new unsatisfiability from 100 to 0. One defective source axiom, one correspondence, one hundred dead operational classes.
3. **Orientation search** (§3), a repair operator in which the search variable is the *direction* of each correspondence rather than its presence. Standard alignment repair removes mappings from a diagnosis; orientation search first weakens $\equiv$ to a subsumption, then attempts completion back toward $\equiv$, with the reasoner refereeing every step against a coherence criterion and a conservativity criterion. Each rejected direction is returned with the reasoner's counterexample, so the artifact documents the disagreement between the two standards rather than hiding it.
4. **Certified bridges** (§4): a 21-axiom IES↔HQDM bridge produced by the search, an IES↔BFO bridge whose four targets were specified by hand and then certified, and their union over the complete three ontologies with direct HQDM↔BFO anchors, all at zero new unsatisfiable classes and zero conservativity violations. An ablation on the BFO bridge (§4.4) separates a mapping that is too strong from one that is aimed at the wrong target, and shows that only the second diagnosis is the right one here.
5. **A systems comparison** (§5). LogMap and LogMapLt, evaluated on the same pairs under the same two criteria, accept the documented false friend, map onto natively unsatisfiable classes, and carry 87 and 214 new unsatisfiable classes plus over a hundred conservativity violations each. Given our expert input, LogMap's repair returns a certifiably safe alignment, but a uniformly reversed one that loses all five provable equivalences. An LLM-oracle probe (6/12 base, 7/12 IES fine-tune) completes the picture: classical and neural systems all fail on the same pairs, the ones where lexical evidence misleads, which is where the reasoner is doing work no other component substitutes for.

All ontologies used are public, all code and certified artifacts are released, and the reasoning is deterministic (see the availability statement at the end of the paper).

2. Materials and background

2.1. The three ontologies

Table 1 reports what we measured on the exact vendored files, not what the documentation claims. Two properties are worth noting immediately. IES declares 511 named classes and *zero* class-disjointness

axioms; it is a rich taxonomy with almost no exclusion structure. HQDM declares 229 named classes and 14 disjointness axioms, introduced by the EXPRESS-to-OWL rendering. BFO 2020 is small (35 named classes) and heavily constrained (19 disjointness axioms), which is exactly why it is the ontology that punishes a careless mapping hardest.

Table 1
The three ontologies as measured on the vendored files. Named classes counts `rdf:type owl:Class` subjects with an IRI. Disjointness counts `owl:disjointWith` triples; none of the three uses `owl:AllDisjointClasses` or `owl:disjointUnionOf`. Unsatisfiable classes are computed by HermiT on each ontology alone.

Ontology	triples	named classes	object props	disjointness	unsat (alone)
IES (`ies-common.ttl`)	4,039	511	162	0	**0**
HQDM (`hqdm.owl`)	3,127	229	74	14	**39**
BFO 2020 (`bfo.owl`)	1,221	35	0	19	**0**

A note on names: the `dstl/IES4` repository was archived in March 2025 (custodianship moved from Dstl to the cross-government IES Working Group under DBT; last public release v4.3.1); the canonical successor is `IES-Org/ont-ies`, whose `ies-common.ttl` we vendor and measure. Counts differ between lineages, so all IES claims here are about the file in Table 1, identified by content hash in the released artifact.

IES and HQDM are both four-dimensionalist: an individual is a spatio-temporal extent, and what a three-dimensionalist calls a state of an object is a temporal part of that extent. BFO is three-dimensionalist at its core, partitioning existence into `bfo:continuant` and `bfo:occurrent` and declaring them disjoint. That single disjointness is the whole difficulty of the IES-to-BFO join, and §4.4 shows it is crossable but only in one place.

2.2. The crosswalk

The input alignment is a hand-curated crosswalk of 17 correspondences in SSSOM [9]: 13 `skos:closeMatch` (10 class pairs, 3 property pairs) and 4 `skos:relatedMatch` recorded as warnings rather than as mappings. It is accompanied by six documented divergences, pairs that look like they map and do not. The headline divergence is a false friend that any label-based matcher gets exactly backwards: `ies:Event` is a durative happening with participants, while `hqdm:event` is an instantaneous temporal boundary with no duration and no participants. The true counterpart of `ies:Event` is `hqdm:activity`. The crosswalk records `ies:Event` $\sim$ `hqdm:event` at confidence 0.25 purely to carry the warning. We return to this pair in §5.3, where both language models we test walk straight into it.

2.3. Consistency, coherence, conservativity

We use three standard criteria and are pedantic about the difference because the central negative result of this paper is that the first one is useless here.

An ontology $\mathcal{O}$ is *consistent* if it has a model, and *coherent* if no named class is unsatisfiable ($\mathcal{O} \not\models C \sqsubseteq \bot$ for named C). Coherence implies consistency; the converse fails, and not exotically: a consistent ontology can be satisfied by a model in which a hundred of its classes are simply empty. Alignment incoherence and its repair are well studied [2, 3, 4], and yet consistency checking remains what deployment pipelines actually run.

The third criterion is *conservativity* [5, 6]. Let $\mathcal{O}_1, \mathcal{O}_2$ be the input ontologies and $\mathcal{M}$ the alignment read as axioms. $\mathcal{M}$ violates conservativity if the merge entails a subsumption between two classes of the *same* input ontology that the input ontology did not itself entail. Such an entailment is a statement about IES that IES never made, produced by the act of aligning it to something else. A consumer who trusts the merged model is now reasoning with axioms no standards body ever wrote.

Our acceptance criteria are therefore:

S1′ (relative coherence). The merge introduces no unsatisfiable class beyond those already unsatisfiable in the inputs alone. We use the relative form because HQDM's 39 native casualties (§4.1) would otherwise make every candidate fail identically and carry no signal.

S2 (conservativity). The merge entails no new subsumption between two classes of the same input ontology.

3. Orientation search

3.1. The operator

Alignment repair conventionally computes a diagnosis, a minimal subset of mappings whose removal restores coherence, and deletes it. Deletion is a blunt move: it throws away a correspondence that a domain expert asserted for good reason, and it records nothing about *why* the correspondence failed.

Orientation search replaces deletion with re-orientation. A curated crosswalk gives us (s, o) pairs but, because SKOS carries no logical commitment, it does not tell us the logical strength of the relation. That strength is precisely what we should be searching over. For each correspondence we consider the lattice

$$\{\, s \equiv o \,\} \;\succ\; \{\, s \sqsubseteq o \,\},\ \{\, s \sqsupseteq o \,\} \;\succ\; \{\, \emptyset \,\}$$

and, when even the forward subsumption fails because the target is defective, a fifth move: *weaken to ancestor*, replacing o by a superclass of o, in the spirit of axiom weakening [7].

The search runs in two phases. **P1** accepts each forward edge $s \sqsubseteq o$, weakening the target upward while the edge fails S1′ or S2. **P2** greedily attempts each reverse edge $o \sqsubseteq s$; where it survives, the pair completes to a full equivalence. The result declares the standards equivalent exactly where they agree and one-directionally linked exactly where they do not.

Algorithm 1 Orientation search

Require: input ontologies $\mathcal{O}_1, \mathcal{O}_2$; correspondences C; reasoner $\mathcal{R}$

$U_{\text{nat}} \leftarrow \mathcal{R}.\textsc{Unsat}(\mathcal{O}_1 \cup \mathcal{O}_2)$ ▷ native baseline
$B \leftarrow \emptyset$
for all $(s, o) \in C$ **do** ▷ P1: forward, weakening to ancestor
  $t \leftarrow o$
  while $t \neq \top$ **and** $\neg\textsc{Certify}(B \cup \{s \sqsubseteq t\})$ **do**
    $t \leftarrow \textsc{Parent}(t)$
  end while
  if $t \neq \top$ **then** $B \leftarrow B \cup \{s \sqsubseteq t\}$
  end if
end for
for all $(s, o) \in C$ with $s \sqsubseteq o \in B$ **do** ▷ P2: completion toward $\equiv$
  if $\textsc{Certify}(B \cup \{o \sqsubseteq s\})$ **then** $B \leftarrow B \cup \{o \sqsubseteq s\}$
  else record $\mathcal{R}$'s witness as the justification for one-directionality
  end if
end for
return B

function $\textsc{Certify}(B)$
  $U \leftarrow \mathcal{R}.\textsc{Unsat}(\mathcal{O}_1 \cup \mathcal{O}_2 \cup B)$
  $V \leftarrow \{\, C_1 \sqsubseteq C_2 \text{ entailed by the merge} \mid C_1, C_2 \in \mathcal{O}_i,\ \mathcal{O}_i \not\models C_1 \sqsubseteq C_2 \,\}$
  return $(U \setminus U_{\text{nat}} = \emptyset) \wedge (V = \emptyset)$ ▷ S1′ and S2
end function

3.2. Instantiation

The reasoner is HermiT [8] via owlready2, called on the complete merged ontologies, not on fragments and not on hand-selected axiom subsets. The run costs on the order of 35 reasoner calls and completes in about two minutes on a laptop. The procedure is deterministic: HermiT is a DL prover with no learned component, so the same inputs give the same certified output.

Two design decisions are worth defending. First, S1′ is relative rather than absolute. If we demanded absolute coherence, every candidate would fail on HQDM's 39 native casualties and the search would return nothing; the relative form isolates the damage the *alignment* does. Second, P2 is greedy rather than exhaustive, so the returned bridge is a maximal certified bridge under a fixed correspondence order rather than a provably maximum one. We consider this an honest limitation (§7) rather than a defect, since every axiom in the output is individually certified regardless of order.

4. Results

4.1. The target ontology is broken before we touch it

Reasoned alone, `hqdm.owl` has **39 unsatisfiable classes** (Table 1). IES alone and BFO alone have none. The casualties are concentrated in HQDM's relationship-derived branch, including `hqdm:association`, `hqdm:employment`, `hqdm:ownership`, `hqdm:participant`, `hqdm:sign`, `hqdm:asset`, `hqdm:transfer_of_ownership` and `hqdm:physical_quantity`. The mechanism is the collision between the EXPRESS relationship heritage, in which a relationship is a first-class entity, and the `hqdm:class`/`hqdm:relationship` and `hqdm:abstract_object`/`hqdm:spatio_temporal_extent` disjointness axioms introduced by the OWL rendering. The file's own header warns of known consistency issues; what we add is the quantification and, more importantly, the identification of *which* classes are affected, because that is what an alignment consumer needs.

Three consequences follow for ontology matching practice. A matcher that maps onto any of those 39 classes has produced a correspondence whose target is provably empty: the mapping is vacuously satisfied and carries no information. A reference alignment built by human curators against this file may contain such mappings, since the defect is invisible in the asserted hierarchy. And an evaluation that reports precision and recall against such a reference is reporting agreement about empty sets. We suggest that a coherence report on each *input* ontology should be a routine part of publishing a matching test case, alongside class and property counts.

4.2. Consistent, and thoroughly incoherent

We promote all 13 `closeMatch` correspondences to `owl:equivalentClass` and `owl:equivalentProperty` and merge the complete ontologies. The result:

Table 2
Naive equivalence promotion of the expert crosswalk. The consistency check, which is what deployment pipelines typically run, passes.

Criterion	Result
Consistency (a model exists)	**passes**
New unsatisfiable classes (beyond the native 39)	**100**
New source-internal subsumptions (S2 violations)	**218**

The 100 casualties are IES operational classes: `ies:Accused`, `ies:Arrested`, `ies:ArrestingOfficer`, `ies:Assessor`, `ies:Witness`, `ies:Customer`, `ies:Authoriser`, `ies:AtWar`. These are the classes an intelligence or policing application actually instantiates. The 218 conservativity violations include `ies:Entity` $\sqsubseteq$ `ies:State`, `ies:Investigation` $\sqsubseteq$ `ies:State`

and ies:OfferForSale ⊑ ies:Entity: statements about IES, in IES's own vocabulary, that IES does not make and that arise solely from having been aligned to HQDM.

Isolation. Exactly one of the ten class targets, hqdm:participant, is among HQDM's 39 native casualties. The certified run (§4.3) weakens exactly one edge, ies:EventParticipant ⊑ hqdm:participant replaced by ies:EventParticipant ⊑ hqdm:state, one hop up the hierarchy, and new unsatisfiability falls from 100 to 0. The single edge into the defective region is therefore what wires IES into HQDM's broken branch, and one correspondence onto one bad class takes down a hundred classes in the other ontology. This is the strongest practical argument we can make for reporting input-ontology coherence: the cost of a single mapping onto an unsatisfiable target is not local.

4.3. The certified IES↔HQDM bridge

Orientation search returns **21 directed axioms** at 0 new unsatisfiable classes and 0 conservativity violations (Table 3). Five class pairs and all three property pairs complete to full equivalence. Four class pairs remain one-directional, each carrying the reasoner's counterexample, and one is weakened.

Table 3
The certified bridge. Prefixes are omitted: every source is ies: and every target hqdm:. The last column is the number of S2 violations the *reverse* edge would have introduced, which is why the row is one-directional. Every verdict is a reasoner decision, not a curator's judgement.

Source (ies:)	Target (hqdm:)	Verdict	S2 if reversed
Element	spatio_temporal_extent	full ≡	0
Thing	thing	full ≡	0
Event	activity	full ≡	0
PossibleWorld	possible_world	full ≡	0
ClassOfElement	class_of_spatio_temporal_extent	full ≡	0
Entity	individual	forward only	104
State	state	forward only	109
PeriodOfTime	period_of_time	forward only	1
ParticularPeriod	period_of_time	forward only	5
EventParticipant	participant → state	weakened	n/a
isPartOf	part_of	full ≡	0
isParticipantIn	participant_in	full ≡	0
isStateOf	temporal_part_of	full ≡	0

The witnesses the reasoner returned for the four rejected reverse edges are the scientific content of this result, so we give them in full. Each violation set is stored sorted, and every witness quoted below is among the first six entries of its sorted set, so the citations are reproducible rather than an artifact of the order in which the reasoner happened to enumerate them:

ies:Entity ∼ hqdm:individual **(104 violations).** The reverse forces ies:Arrest ⊑ ies:Entity, and likewise ies:AccountAdminEvent, ies:AgreementExecution, ies:Assessment and 100 further IES classes. HQDM asserts hqdm:activity ⊑ hqdm:individual, while IES holds ies:Event and ies:Entity apart, so every IES happening is dragged under ies:Entity. HQDM's hqdm:individual is strictly broader than IES's notion of a whole-life persisting thing.

ies:State ∼ hqdm:state **(109 violations).** The reverse forces ies:Arrest ⊑ ies:State by the same route, since HQDM's hqdm:individual sits under hqdm:state. An arrest becomes a state of IES.

`ies:PeriodOfTime` $\sim$ `hqdm:period_of_time` **(1 violation).** The reverse forces `ies:PossibleWorld` $\sqsubseteq$ `ies:PeriodOfTime`, because HQDM asserts `hqdm:possible_world` $\sqsubseteq$ `hqdm:period_of_time` and IES does not.

`ies:ParticularPeriod` $\sim$ `hqdm:period_of_time` **(5 violations).** The reverse forces `ies:PeriodOfTime` $\sqsubseteq$ `ies:ParticularPeriod`, collapsing the distinction between a period and an ISO 8601-anchored particular one, and likewise for `ies:ArbitraryPeriod`, `ies:RecurringPeriod` and `ies:PossibleWorld`.

These are not failures of the crosswalk. They are the two standards disagreeing, localised to a named class by a prover. The `ies:Entity` $\sim$ `hqdm:individual` case says something a human curator can act on: HQDM subsumes activities under individuals, IES insists an event is not an entity, so the two notions of whole-life persisting thing are genuinely different and the correct relation is a subsumption, not an identity. That is a finding about the standards produced as a by-product of certifying the alignment, and it is exactly the kind of content that an alignment published as a bare similarity score cannot carry.

We note the shape of the outcome. Where the two 4D models genuinely agree, the bridge is full equivalence. Where they disagree, it is exactly as weak as the disagreement requires and no weaker. A diagnosis-and-delete repair would have returned a smaller alignment with less information in it.

4.4. Crossing to BFO: only via the occurrent branch

BFO is where the 4D/3D mismatch has to be paid for. Unlike the HQDM case, the four BFO targets were specified by hand and then certified; orientation search verified and oriented them but did not discover them, and §7 explains why weakening alone could not have.

The bridge is

`ies:Element` $\sqsubseteq$ `bfo:occurrent` `ies:Entity` $\sqsubseteq$ `bfo:history`
`ies:Event` $\sqsubseteq$ `bfo:process` `ies:State` $\sqsubseteq$ `bfo:occurrent`
`cwb:hasHistory` : `bfo:material entity` $\rightarrow$ `bfo:history` (declared by the bridge)

certified at 0 new unsatisfiable classes and 0 conservativity violations. The one axiom in it that a curator would not write is `ies:Entity` $\sqsubseteq$ `bfo:history`. The obvious choice is `ies:Entity` $\equiv$ `bfo:material entity`, and Table 4 shows what that choice costs by changing *only* that axiom and holding the other three fixed.

Table 4
The cost of one axiom. Every row merges the complete IES and BFO with `ies:Element` $\sqsubseteq$ `bfo:occurrent`, `ies:Event` $\sqsubseteq$ `bfo:process` and `ies:State` $\sqsubseteq$ `bfo:occurrent` held fixed, and varies only how `ies:Entity` is mapped. Counts are IES classes made unsatisfiable.

Mapping for `ies:Entity`	Rest of bridge	IES classes collapsed
$\equiv$ `bfo:material entity`	absent	0
$\sqsubseteq$ `bfo:material entity`	absent	0
$\equiv$ `bfo:material entity`	present	**101**
$\sqsubseteq$ `bfo:material entity`	present	**100**
$\sqsubseteq$ `bfo:history`	present	**0**

Three things follow. First, the collapse is a property of the bridge as a whole, not of one mapping in isolation: aligning `ies:Entity` to `bfo:material entity` on its own is harmless, because IES declares no disjointness and nothing then connects it to BFO's occurrent branch. Only once the rest of the bridge puts IES's 4D backbone on that branch does the `bfo:continuant`/`bfo:occurrent` disjointness fire and empty 101 IES classes (`ies:Account`, `ies:Actor`, `ies:Aircraft`, `ies:Bank`, …). An alignment cannot be certified one correspondence at a time.

Second, weakening does not rescue it. Replacing the equivalence with a plain subsumption still kills 100 classes. This is the case that shows why the weaken-to-ancestor operator of §3 is not sufficient on its own: walking up from `bfo:material entity` reaches `bfo:independent continuant`, `bfo:continuant` and finally `bfo:entity`, at which point the axiom is safe but vacuous. The ancestor chain never passes through `bfo:history`.

Third, re-targeting does rescue it, completely. BFO 2020 already ships the class a 4D standard is talking about: `bfo:history` (`BFO_0000182`), a process that is the totality of what happens to a material entity, sitting natively on the occurrent branch. Changing that single axiom takes the count from 101 to 0. The naive mapping was not too strong; it was aimed at the wrong target, and the distinction matters because only one of those two diagnoses is fixed by weakening. A 4D `ies:Entity` is not a BFO continuant; it is the continuant's history. The bridge declares `cwb:hasHistory` so that a consumer can still get from the continuant to the 4D object, which is the relation the naive equivalence was groping for.

4.5. All three in one artifact

The IES↔HQDM and IES↔BFO bridges are not merely individually safe. Their union over the complete IES ∪ HQDM ∪ BFO is also certified at 0 new unsatisfiable classes and 0 conservativity violations, and it admits direct HQDM↔BFO anchors on top, found the same reasoner-refereed way: `hqdm:spatio_temporal_extent` ≡ `bfo:occurrent`, `hqdm:individual` ≡ `bfo:history`, `hqdm:activity` ⊑ `bfo:process`, `hqdm:state` ⊑ `bfo:occurrent`, `bfo:temporal region` ⊑ `hqdm:period_of_time`. The fused artifact is 32 subsumption axioms plus the `cwb:hasHistory` declaration. `hqdm:participant` is deliberately excluded from the HQDM↔BFO anchors: it is one of the 39 native casualties, so any mapping onto it would be vacuously true.

That the two independently certified bridges compose without new violations is not guaranteed by their individual certificates; we checked it rather than assumed it. Composition of certified alignments must, in general, be re-certified.

5. Do existing systems produce this bridge?

Mature systems exist for both halves of this problem: matchers that propose correspondences, and repair facilities that restore coherence. We ran the natural incumbents on exactly this material and evaluated every output with the same HermiT harness and criteria. Three questions: does a state-of-the-art matcher *find* the expert backbone; is its output *safe* to consume as logic; and, given the expert crosswalk, does an established *repair* facility recover what orientation search recovers?

5.1. Classical matchers on the pair

We ran LogMap and LogMapLt (the July-2021 standalone distribution, `logmap-matcher-4.0.jar`, default configuration) on the vendored IES–HQDM pair, and LogMap on IES–BFO. Table 5 evaluates each output alignment, read as axioms per the system's own asserted relations, against S1′ and S2.

Three observations, each of which independently justifies a reasoner in the loop.

First, **both matchers accept the false friend** `ies:Event` ≡ `hqdm:event`, and neither finds `ies:Event` ∼ `hqdm:activity`. The consequence is not merely a wrong pair: read as logic, LogMap's output entails `hqdm:activity` ⊑ `hqdm:event` *inside HQDM* (every durative activity becomes an instantaneous boundary): an S2 violation against an ontology the matcher was supposed to align, not rewrite. Both matchers also accept mappings onto `hqdm:asset` and `hqdm:agreement_execution`, two of the 39 natively unsatisfiable classes (vacuously true correspondences), and both map RDF(S) vocabulary itself (`rdfs:Class` ≡ `hqdm:class`; LogMapLt adds `rdfs:subClassOf` ≡ `hqdm:causes_by_class`).

Second, **the discovery problem on this pair is genuinely hard, which is the point of offering it as a test case**: the backbone correspondences are exactly the ones lexical evidence cannot find. LogMap recovers 3 of 10; on IES–BFO it recovers none, its only substantive mapping being

Table 5
Matcher outputs on the vendored pairs, evaluated with HermiT under the same criteria as the certified bridge. "Backbone" scores correct target / wrong target / missed against the pair's gold targets: the ten expert class correspondences for IES–HQDM, the four certified bridge targets for IES–BFO. "Unsat targets" are accepted mappings whose HQDM target is one of the 39 natively unsatisfiable classes. The certified bridge row is the output of §4.3.

System (pair)	mappings	backbone	false friend	unsat targets	new unsat	S2 viol.
LogMap (IES–HQDM)	24	3/1/6	**yes**	2	**87**	**103**
LogMapLt (IES–HQDM)	36	3/1/6	**yes**	2	**214**	**112**
LogMap (IES–BFO)	2	0/1/3	—	0	0	**2**
certified bridge (IES–HQDM)	13	10/0/0	no	0	**0**	**0**

`ies:Entity` ≡ `bfo:entity`, the same bare root-label match both LLMs produce in §5.3; even two mappings suffice for two conservativity violations (`ies:ObjectName` ⊑ `ies:Entity`: a name becomes an entity). Matching from scratch is harder than certifying an expert crosswalk, and we do not claim to out-match matchers; the like-for-like comparison is repair, next.

Third, **LogMap's internal repair is incomplete by design**: a scalable propositional projection rather than full DL reasoning [4]. This pair shows the residue concretely, in the 87 new unsatisfiable classes that survive it. That is not a defect of LogMap so much as a measurement of exactly the gap a complete reasoner in the certification loop closes, a gap the OAEI's DISO track design already anticipates by shipping repaired and unrepaired reference alignments side by side.

5.2. LogMap's repair facility versus orientation search

The like-for-like comparison: we gave LogMap's DEBUGGER facility the same input orientation search received (the ten expert class correspondences promoted to equivalences, over the complete IES and HQDM), with module extraction and its post-repair HermiT check enabled.

The outcome is instructive on both sides. LogMap keeps all ten mappings and its output is *certifiably safe*: our harness confirms 0 new unsatisfiable classes and 0 conservativity violations. Its repair, however, achieves safety by a single uniform move: every equivalence is weakened to the same direction, `hqdm:X` ⊑ `ies:Y`, discarding the entire forward (IES-to-HQDM) content of the crosswalk. Concretely, against the certified bridge of Table 3:

- **All five provable equivalences are lost.** Orientation search proves `ies:Element` ≡ `hqdm:spatio_temporal_extent` and four more; LogMap keeps only the reverse half of each.
- **On the four divergent pairs, LogMap keeps the semantically wrong direction.** `hqdm:individual` ⊑ `ies:Entity` asserts that every HQDM individual, including its activities, is an IES entity, precisely what IES denies by separating `ies:Event` from `ies:Entity`. The assertion is undetectable only because IES declares no disjointness axioms, and it survives S2 solely because LogMap simultaneously dropped the forward edges that would have exposed it. Orientation search, which preserves the expert's forward commitment as its baseline, rejects exactly this direction with a named counterexample.
- **The mapping onto the broken class is kept, unflagged.** `hqdm:participant` ⊑ `ies:EventParticipant` subsumes a provably empty class; coherence checking cannot distinguish vacuous from safe. Orientation search instead weakens the target one hop to `hqdm:state` and keeps a contentful axiom.

The comparison is therefore not repair-fails-versus-repair-succeeds; both outputs are coherent and conservative. It is *retained logical content at equal safety*: 10 uniformly-reversed axioms against 21 directed axioms with five class and three property equivalences, the expert's forward direction wherever safe, and a counterexample attached to every refusal. A criterion that checks only coherence stops at the first safe answer; adding conservativity and completion-toward-equivalence finds the strongest one.

5.3. An LLM oracle

LLM-as-oracle is now a standard component of matching systems, used to adjudicate candidate correspondences. We wanted a direct test on this alignment, using our certified bridge as gold, in the shape of the DISO-OAEI ranking task: given a source class and a candidate pool from the target ontology, pick the best target or return NIL.

The benchmark is 12 queries (8 into HQDM, 4 into BFO) built from the certified bridge. We evaluate two systems: `Qwen3-Coder-30B-A3B-Instruct` at 8-bit, and an IES-specialised fine-tune of the same base model, both served locally at temperature 0. Results in Table 6.

Table 6
Mapping-oracle accuracy against the HermiT-certified bridge. $n = 12$.

System	into HQDM (8)	into BFO (4)	total
Qwen3-Coder-30B-A3B-Instruct (8-bit)	5	1	**6/12**
+ IES fine-tune	6	1	**7/12**

The aggregate is uninformative at this sample size and we do not claim a significant difference between the two systems. The *error structure* is the result worth reporting, and it is consistent across both:

- **Both models pick `hqdm:event` for `ies:Event`.** This is the false friend the crosswalk documents as its single most dangerous correspondence: a durative happening with participants mapped onto an instantaneous temporal boundary. The lexical pull of the identical label defeats both models, and the fine-tune does not help, because the trap is not a vocabulary problem.
- **Neither model reaches `bfo:history`.** Asked for the BFO counterpart of `ies:Entity`, both answer `bfo:entity`: a shallow label match onto BFO's root. The certified answer requires knowing that a 4D whole-life individual is an occurrent, which is a commitment neither model retrieves.
- **Both succeed where the mapping is easy.** `hqdm:thing`, `hqdm:state`, `hqdm:period_of_time`, `hqdm:possible_world` and `hqdm:spatio_temporal_extent` are all recovered by both.

The pattern is that an LLM oracle is accurate exactly where a lexical baseline would also be accurate, and fails exactly on the pairs where the two standards genuinely diverge, which is where the alignment is hard and where the value lies. On this evidence, an LLM is a reasonable candidate generator and a poor adjudicator of upper-ontology disagreement; the reasoner is doing work the model cannot substitute for. We report this as a probe with an explicit sample size, not as a benchmark, and we say what would make it one in §7.

6. Discussion

Publish the coherence report with the test case. The 39-class defect in `hqdm.owl` was invisible to everything except a DL reasoner run on the input alone; it cost nothing to find and changes how any alignment into HQDM should be read. A matching test case should ship a coherence report on each input, and a reference alignment should flag any correspondence whose target is unsatisfiable: precision and recall over such correspondences measure agreement about the empty set.

Direction is information, and alignment formats should carry it. A correspondence published as $\langle s, o, 0.85 \rangle$ discards the distinction between "these are the same" and "every s is an o but not conversely, and here is the individual that shows it." The second is more useful and, as §4.3 shows, often the true relation between standards built by different communities. SSSOM and EDOAL can express directed correspondences; what is missing is the habit of computing the direction rather than defaulting to equivalence. Orientation search is one way to compute it.

Relevance to OAEI. The DISO track introduced in the OAEI 2026 campaign evaluates alignment over defence, intelligence and security ontologies, and ships both a repaired and an unrepaired reference alignment ($R^{+}_{\approx}$ and $R_{\approx}$), which makes coherence an explicit first-class concern of the evaluation rather than a post-processing step. This paper's material has exactly that two-reference structure, produced by the method itself: the naively promoted crosswalk of §4.2 is an unrepaired reference (consistent, incoherent, 218 conservativity violations) and the certified bridge of §4.3 is its repaired counterpart, with the delta between them fully explained by machine-found counterexamples. The DISO collection ships IES, BFO and JC3IEDM but not HQDM, so the IES–HQDM leg is exactly the join the collection cannot currently evaluate, and the pair has a profile a track wants: small enough for exhaustive reasoner-based evaluation, yet adversarial for lexical matchers, with an exact-label false friend as its hardest correspondence and a correct BFO target sharing no vocabulary with its source (§5.1 measures how adversarial; DISO distributes IES v5, so a contributed case would pin its lineage as §2 does). We offer the three ontologies, the expert crosswalk with its six documented divergences, and both reference alignments as a candidate test case with one property the current track does not yet score: the repaired reference is *directed*, each non-equivalence carrying its justifying counterexample, so a system can be scored not only on which pairs it finds but on whether it gets the logical strength of each pair right.

7. Limitations

The crosswalk is a 17-correspondence backbone over two upper ontologies, not a full alignment; above the 4D backbone the powertype hierarchies of the two standards stop being parallel and become a genuine matching problem that we have not attempted here. The correspondences were hand-curated by a single author and their confidences are the curator's subjective strength rather than a measured score; what is certified is the *logical form* of each correspondence, not its correctness as a claim about the domain. Orientation search's completion phase is greedy under a fixed correspondence order, so the returned bridge is maximal rather than provably maximum, and P1's weakening walks a single ancestor chain rather than searching the full weakening lattice.

The most important limitation is one the BFO ablation makes concrete. Weakening is confined to the ancestors of the asserted target, so the operator can never reach a correct target that lies on a different branch. §4.4 is exactly that case: from `bfo:material entity` the ancestor chain leads only to `bfo:entity`, where the axiom is safe but says nothing, and `bfo:history` is unreachable. We supplied the four BFO targets by hand and used the search to certify and orient them. Making re-targeting part of the search, rather than part of the curator's judgement, is the obvious next step and we have not taken it; a candidate generator over the target ontology, filtered by the same two criteria, is the natural shape.

The systems comparison has its own fairness bounds: the July-2021 LogMap standalone in default configuration (tuned or interactive use might do better); the DEBUGGER fed the ten class correspondences only, not the three property mappings; AgreementMakerLight [21] and BERTMap [10] not run (older-JVM release line; per-pair language-model fine-tuning). Extending to them, and to the conservativity-repair algorithms of [5], is the natural next experiment.

The LLM probe is 12 queries on two models from one family, which supports the qualitative claim about error structure and supports no quantitative claim at all; a proper version would cover multiple model families, multiple prompt formats, and a candidate pool large enough for Hits@k to be meaningful. Finally, we certify against HermiT alone. Every result was re-checked by invoking the HermiT command-line tool directly on the serialised merges, bypassing owlready2's object model, and it agrees exactly (IES alone 0, HQDM alone 39, IES+HQDM with the certified bridge 39 and the identical set, IES+BFO with the certified bridge 0). That is a check on the harness, not an independent prover; agreement with a second reasoner would strengthen the certificates.

8. Related work

Alignment incoherence and its repair are the direct ancestors of this work. Meilicke's analysis of alignment incoherence and the diagnosis-based repair tradition [2, 3] established that a coherent alignment is the object of interest and that repair is a minimal-hitting-set problem over mappings. LogMap [4] made scalable logic-based repair a practical component of a matching system: its repair module identifies mappings implicated in unsatisfiability and discards or weakens confidence in them during matching. The conservativity principle and its violation-repair algorithms [5] supplied the second criterion we use, and the theory of conservative extensions and safe ontology reuse [6] is where that criterion comes from. Our contribution relative to this line is narrow and, we think, useful: repair here happens *after* an expert has committed to the correspondence set, so removal would discard curated knowledge, and the free variable that remains is the *direction* of each correspondence. Making direction the repair variable, with completion toward equivalence and a counterexample attached to every refusal, turns the repair log into documentation of where two government standards disagree, arguably the more valuable artifact.

Axiom weakening [7] is the other parent, and Li and Lambrix [20] developed weakening-and-completing for EL ontologies; our P1+P2 applies that shape at the alignment layer, to the mapping target rather than an axiom of either input.

On the machine-learning side, BERTMap [10] established the pattern of a learned matcher followed by a symbolic repair stage, and recent work applies LLMs as candidate generators or oracles over ontology matching [11]. Our §5.3 probe is a small negative datapoint for the oracle role precisely where lexical evidence misleads. MELT [12] is the evaluation platform on which the OAEI 2026 campaign runs, including the DISO track that motivates §6.

On the ontologies themselves, HQDM derives from West's data-modelling work [13] in the BORO tradition [14]; BFO is described in [15] and standardised as [16]. A companion census of public ontologies [17] reported that neither IES nor HQDM declares exclusion structure adequate for counterfactual reasoning; this paper asks the complementary question of what their declared axioms do to each other when joined.

On IES–BFO specifically, Bailey et al. [19], a team that includes the original authors of IES, BORO and BFO, compare the two frameworks through parallel modelling of competency-question scenarios and propose graph-level mapping patterns, including a pattern relating an IES4 4D individual to a BFO material entity *together with its* `bfo:history`. Our certified bridge sharpens that insight: `ies:Entity` subsumes safely under `bfo:history` but not `bfo:material entity` (Table 4); of their pattern's two legs, exactly one survives as an axiom. They explicitly defer axiomatic, provable alignment to future work; the present paper is that step for the 4D backbone. We previously released a SKOS-level (not reasoner-checked) sketch crosswalk from JC3IEDM to IES [18]; this is, to our knowledge, the first *certified* alignment involving IES and the first public IES–HQDM alignment. Euzenat and Shvaiko [1] remains the reference text.

9. Conclusion

Two upper ontologies from the same intellectual tradition, aligned by an expert on 17 backbone correspondences, produce a merge that passes a consistency check and kills a hundred operational classes; one of them was already killing 39 of its own before anyone aligned anything. Both facts are cheap to find with a reasoner, invisible without one, and absent from any similarity score.

The repair is not deletion. Treating the direction of each correspondence as the search variable, with a prover refereeing coherence and conservativity at every step, returns a bridge that keeps every expert mapping, states full equivalence wherever the standards agree, and attaches a machine-found counterexample wherever they do not. No system we tested recovers that bridge: the matchers fall for the false friend, and the repair that succeeds discards every provable equivalence. BFO's 3D core is crossable only through the occurrent branch: a 4D entity is a continuant's history, not the continuant.

Declaration on Generative AI

During the preparation of this work, the author used Anthropic Claude in order to: draft and revise text, assist literature triage, and prepare LaTeX. The locally served Qwen3-Coder-30B models in Section 5.3 are systems under evaluation, not writing tools. All experiments, reported numbers and reasoner runs were designed, executed and verified by the author. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

A. Data and code availability

The crosswalk (SSSOM and SKOS/PROV-O), divergences, SHACL shapes, orientation search, the three certified bridges with full decision/counterexample logs, the LogMap/LogMapLt outputs with their evaluation harness (`reasoning/systems/`, `systems_eval.py`), and the LLM benchmark are released under CC-BY-4.0 at https://github.com/fabio-rovai/ies-hqdm-crosswalk; everything reproduces with `python reasoning/orientation_search.py` given any JDK. The input ontologies are public (IES: OGL v3; HQDM: Apache-2.0, Crown Copyright; BFO 2020: CC-BY). This work is independent and not affiliated with or endorsed by Dstl, GCHQ, the IES Working Group or the NDTP.

References

[1] J. Euzenat and P. Shvaiko. *Ontology Matching*, 2nd edition. Springer, Heidelberg, 2013.

[2] C. Meilicke. *Alignment Incoherence in Ontology Matching*. PhD thesis, University of Mannheim, 2011.

[3] C. Meilicke and H. Stuckenschmidt. An efficient method for computing alignment diagnoses. In *Proc. 3rd International Conference on Web Reasoning and Rule Systems (RR)*, LNCS 5837, pages 182–196, 2009.

[4] E. Jiménez-Ruiz and B. Cuenca Grau. LogMap: logic-based and scalable ontology matching. In *Proc. 10th International Semantic Web Conference (ISWC)*, LNCS 7031, pages 273–288, 2011.

[5] A. Solimando, E. Jiménez-Ruiz, and G. Guerrini. Detecting and correcting conservativity principle violations in ontology-to-ontology mappings. In *Proc. 13th International Semantic Web Conference (ISWC)*, LNCS 8797, pages 1–16, 2014.

[6] B. Cuenca Grau, I. Horrocks, Y. Kazakov, and U. Sattler. Modular reuse of ontologies: theory and practice. *Journal of Artificial Intelligence Research*, 31:273–318, 2008.

[7] N. Troquard, R. Confalonieri, P. Galliani, R. Peñaloza, D. Porello, and O. Kutz. Repairing ontologies via axiom weakening. In *Proc. 32nd AAAI Conference on Artificial Intelligence*, pages 1981–1988, 2018.

[8] B. Glimm, I. Horrocks, B. Motik, G. Stoilos, and Z. Wang. HermiT: an OWL 2 reasoner. *Journal of Automated Reasoning*, 53(3):245–269, 2014.

[9] N. Matentzoglu et al. A simple standard for sharing ontological mappings (SSSOM). *Database*, 2022:baac035, 2022.

[10] Y. He, J. Chen, D. Antonyrajah, and I. Horrocks. BERTMap: a BERT-based ontology alignment system. In *Proc. 36th AAAI Conference on Artificial Intelligence*, pages 5684–5691, 2022. arXiv:2112.02682.

[11] H. B. Giglou, J. D'Souza, F. Engel, and S. Auer. LLMs4OM: matching ontologies with large language models. arXiv:2404.10317, 2024.

[12] S. Hertling, J. Portisch, and H. Paulheim. MELT: matching evaluation toolkit. In *Proc. 15th International Conference on Semantic Systems (SEMANTiCS)*, LNCS 11702, pages 231–245, 2019.

[13] M. West. *Developing High Quality Data Models*. Morgan Kaufmann, 2011.

[14] C. Partridge. *Business Objects: Re-Engineering for Re-Use*, 2nd edition. The BORO Centre, 2005.

[15] R. Arp, B. Smith, and A. D. Spear. *Building Ontologies with Basic Formal Ontology*. MIT Press, 2015.

[16] International Organization for Standardization. *ISO/IEC 21838-2:2021, Information technology — Top-level ontologies (TLO) — Part 2: Basic Formal Ontology (BFO).* ISO, Geneva, 2021.

[17] F. Rovai. Is the semantic web counterfactual-ready? A tractability census of public ontologies. Preprint, 2026.

[18] F. Rovai. JC3IEDM–IES4 crosswalk (sketch). Open Ontologies case study, https://github.com/fabio-rovai/open-ontologies, 2026.

[19] I. Bailey, J. Beverley, H. Blackmore, A. Cola, P. Cripps, G. De Colle, F. Donato, A. Hicks, D. Limbaugh, E. Milivinti, C. Partridge, R. Rafferty, and B. Smith. Comparing Information Exchange Standard and Basic Formal Ontology design patterns. In *Proc. Joint Ontology Workshops (JOWO), FOUST track, co-located with FOIS 2025*, CEUR Workshop Proceedings, Vol. 4176, 2025. https://ceur-ws.org/Vol-4176/foust-6.pdf.

[20] Y. Li and P. Lambrix. Repairing $\mathcal{EL}$ ontologies using weakening and completing. In *Proc. 20th Extended Semantic Web Conference (ESWC)*, 2023. doi:10.1007/978-3-031-33455-9_18. arXiv:2208.00486.

[21] D. Faria, C. Pesquita, E. Santos, M. Palmonari, I. F. Cruz, and F. M. Couto. The AgreementMakerLight ontology matching system. In *Proc. OTM Confederated Conferences (ODBASE)*, LNCS 8185, pages 527–541, 2013.