Enabling Semantic Traceability in Health Data: The Health-RI Semantic Interoperability Initiative

Pedro Paulo F. Barcelos[1,*,†], Niek van Ulzen[1,†], Reinier Groeneveld[1,†], Ana Konrad[1,†], Qasim Khalid[2,†], Shuxin Zhang[3,†], Annemarie Trompert[1,†], and Janet Vos[1,†]

[1] *Health-RI, Jaarbeursplein 6, 3521 AL, Utrecht, The Netherlands*
[2] *Leiden University Medical Center (LUMC), Albinusdreef 2, 2333 ZA Leiden, The Netherlands*
[3] *Amsterdam University Medical Center, Meibergdreef 9, 1105 AZ, Amsterdam-Zuidoost, The Netherlands*

Abstract

This paper presents the Health-RI Semantic Interoperability Initiative, a model-driven, ontology-based framework for FAIR-aligned semantic interoperability in the health and life sciences, grounded in semantic traceability. The Initiative addresses the technically complex, time-consuming, and error-prone nature of manual, case-by-case mappings across standards such as FHIR, OMOP, and openEHR, as well as across the heterogeneous artifacts that use or combine them, without requiring replacement of existing standards or local schemas. It introduces the Health-RI Ontology (HRIO), a common semantic reference model specified in OntoUML as the Computation Independent Model and implemented as a gUFO-based OWL ontology, providing a machine-processable semantic hub. Each external artifact is intended to be aligned to this hub once, rather than through multiple pairwise mappings. To align external artifacts to HRIO, the Health-RI Mapping Vocabulary (HRIV) defines intentional (definitional) meaning-mapping relations that explicitly capture ontological commitments. An illustrative example centered on Person's sex- and gender-related specializations demonstrates how the approach can make distinct conceptualizations explicit and traceable across layers, supporting the mitigation of false agreement when integrating data across systems. The Initiative publishes its artifacts with persistent identifiers and documentation to support reuse and extension.

Keywords

Semantic interoperability, semantic traceability, semantics, health data, FAIR principles, OntoUML, gUFO

1. Introduction

The health and life sciences domain is shaped by many standards and associated data models that guide database design and system implementation. They are applied in diverse contexts, from hospital records to research infrastructures. To enable effective data reuse, the data within these systems often need to be exchanged, compared, and integrated [1], [2]. However, achieving this is technically complex and resource-intensive, particularly when relying on manual, case-by-case mappings between heterogeneous standards and local schemas [3], [4], [5].

In the past, several initiatives have sought to address this challenge, and while they brought valuable contributions, some unintentionally added to the creation of yet another model in an already crowded landscape [6], [7]. Important efforts such as the Fast Healthcare Interoperability Resources (FHIR) [8], the Open Electronic Health Record (openEHR) [9], Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) [10], the Observational Medical Outcomes Partnership (OMOP) [11], and convergence initiatives between them (e.g., [12], [13], [14], [15], [16]) remain

---

SWAT4HCLS 2026: Semantic Web Applications and Tools for Health Care and Life Sciences, March 23–26, 2026, Amsterdam, The Netherlands

[*] Corresponding author.

[†] These authors contributed equally.

pedro.barcelos@health-ri.nl (P. P. F. Barcelos); niek.vanulzen@health-ri.nl (N. van Ulzen); reinier.groeneveld@health-ri.nl (R. Groeneveld); ana.konrad@health-ri.nl (A. Konrad); q.khalid@lumc.nl (Q. Khalid); s.x.zhang@amsterdamumc.nl (S. Zhang); annemarie.trompert@health-ri.nl (A. Trompert); janet.vos@health-ri.nl (J. Vos)

0000-0003-2736-7817 (P. P. F. Barcelos); 0000-0003-0771-3516 (N. van Ulzen); 0009-0006-1783-9082 (R. Groeneveld); 0009-0002-3089-9558 (A. Konrad); 0000-0002-7944-2594 (Q. Khalid); 0000-0003-4715-9070 (S. Zhang); 0009-0000-0967-7294 (A. Trompert); 0000-0001-8802-7002 (J. Vos)

crucial endeavors for interoperability. The framework described in this paper builds directly on these efforts while taking a distinct perspective.

Semantic interoperability refers to the ability of different computer systems to exchange data with unambiguous, shared meaning, so that information retains its intended interpretation across systems. This capability is vital across sectors, and particularly in the health and life sciences, where meaningful reuse of data underpins clinical decision-making, research, and policy-making [3], [17]. Unlike syntactic interoperability, which focuses on the format of data exchange, semantic interoperability ensures that the meaning of the data is preserved and understood consistently across different systems [17], [18].

A critical challenge that semantic interoperability aims to address is the *false agreement* problem, as defined by Nicola Guarino [19]. This problem arises when different agents (e.g., systems or people) use the same term or data structure but implicitly commit to different conceptualizations. Crucially, in such cases, each agent still regards the exchanged information as valid with respect to its own schema or space of allowed interpretations, so that an apparent agreement at the level of data obscures a genuine semantic misalignment [19], [20].

A common approach to address this challenge is to link data elements in local schemas to a shared vocabulary or ontology, so that machines can interpret and process information in a consistent way [21]. In this paper, "data element" refers to a unit of data with a defined meaning and representation, typically implemented as a field or variable within a specific schema. However, many widely used representation formalisms—including Semantic Web ones like OWL—are ontologically underdetermining with respect to domain-level commitments [22] (unless those commitments are made explicit and precise, e.g., through foundationally grounded conceptual modeling [20], [23]). In such languages, similar or even identical formal patterns may be interpreted under different conceptualizations, directly enabling false agreement [18], [22]. As a result, neither similarity nor difference in formal structure is sufficient to establish shared meaning [22].

To prevent this, semantic interoperability frameworks must make their commitments explicit, ensuring that shared terms reflect shared understanding [18], [20], [22]. Building on this, Guizzardi emphasizes that every information system reflects a particular view of reality (i.e., a conceptualization) and that interoperability succeeds only when these commitments are clarified and harmonized across systems [18], [22]. In the context of the FAIR (Findable, Accessible, Interoperable, and Reusable) principles [1], this implies that interoperability must be understood not merely as a technical property, but as a commitment to meaningful, semantically grounded data integration [18].

In practice, however, data producers and consumers usually operate in a landscape where each institution maintains its own data schemas and repositories, often resulting in fragmented "data silos" [4], [17], [24]. These schemas often vary significantly in structure, terminology, and the standards or assumptions they follow [3], [4], [5], [17]. Some schemas adhere to formal standards, while others are locally defined by data authors and vary widely in modeling language, degree of formalization, and level of abstraction. A single schema may be influenced by multiple standards, each introducing different terminologies or modeling assumptions, and documentation is frequently incomplete or ambiguous [4], [5]. This lack of clear, shared semantics introduces significant barriers to data integration and analysis: it increases the risk of misinterpretation, reduces confidence in the data, and ultimately limits reuse across the health data ecosystem [1], [3], [17].

To address these challenges and ensure semantic alignment across the health and life sciences community and institutions, the Health-RI Semantic Interoperability Initiative proposes a model-driven, ontology-based architecture grounded in a foundational ontology and centered on a common semantic reference model. Rather than relying on isolated and potentially ambiguous implementation artifacts, the Initiative is developing a unifying ontology, the Health-RI Ontology. Its goal is to clarify and make explicit the precise meanings of key concepts, focusing on those already present in widely used standards such as FHIR, OMOP, and openEHR. This ontology is designed to serve as a semantic hub that, rather than asserting a single authoritative definition of each concept, makes explicit the different interpretations adopted in external standards and local schemas and allows these artifacts to map their concepts to the interpretations they adopt. As a result, intended meanings become clear and explicit to anyone who needs to interoperate across systems.

A key design choice is that the framework does not require replacing existing standards and local schemas. Research groups and institutions have invested substantial effort and resources in their current solutions, and discarding them is neither realistic nor desirable. Instead, the Initiative is intended to reduce reliance on multiple pairwise mappings between those standards by promoting a single, reusable mapping from each standard or local schema to the Health-RI Ontology (HRIO), which serves as the common semantic reference model. Here, "single, reusable mapping" should be understood per source artifact and version, with mappings curated and revised as standards and local models evolve. For example, once a standard-specific concept (e.g., a FHIR concept) is meaning-mapped to HRIO for a particular release, the same alignment can, in practice, be reused across projects that adopt that same source basis, rather than being recreated in each integration scenario.

This makes the framework non-intrusive: the ontology complements existing efforts and provides a semantic backbone to support interoperability. It can be used as a single semantic hub through which users discover and reuse relations among mapped standards and schemas, including via querying and reasoning over HRIO, so that once a concept or schema element is aligned, relationships to other already-aligned elements become assessable through the hub, thereby avoiding duplicated case-by-case mapping work across integration scenarios. When a new local schema introduces a concept or schema element, an additional meaning mapping is still required, but once it is linked to HRIO, its relationships to other already-mapped elements can be assessed in the same way. By relying on solid theoretical foundations and widely adopted Semantic Web standards and technologies, and, importantly, by engaging with the community, the Initiative aims to make this framework a practical means of achieving semantic interoperability in line with FAIR principles.

The common semantic reference model is expressed in OntoUML, a highly expressive, ontologically well-founded conceptual modeling language grounded in the Unified Foundational Ontology (UFO) [20], [23], and is implemented as a corresponding OWL ontology using gUFO ("gentle" UFO), a lightweight OWL implementation of UFO [25]. To support the alignment of external standards, schemas, and ontologies to the Health-RI Ontology, the Initiative also defines a dedicated mapping vocabulary in OWL.

The remainder of this paper is organized as follows. Section 2 reviews related work on semantic interoperability in health data. Section 3 introduces the Health-RI Semantic Interoperability Initiative's architecture and artifacts, describing its conceptual foundations, ontologies, and the mapping vocabulary. Section 4 illustrates how these elements work together through an illustrative example. Section 5 presents final discussions and future work.

2. Related Work on Semantic Interoperability in Health Data

Recent efforts to operationalize semantic interoperability in health data have largely focused on engineering mappings and transformation mechanisms between widely adopted standards such as FHIR, OMOP, and openEHR. Several frameworks automate this process. Xiao et al.'s FHIR-Ontop-OMOP leverages Ontology-Based Data Access (OBDA)-style mappings to expose OMOP-based repositories as virtual clinical knowledge graphs conforming to the FHIR RDF specification [26]. Kohler et al.'s Eos/OMOP Conversion Language (OMOCL) introduces a declarative mapping language and Extract, Transform, Load (ETL) tool that automate the transformation of openEHR archetypes into OMOP tables [12]. FHIRconnect provides a domain-specific language and execution engine to support bidirectional, archetype-based mappings between openEHR and FHIR [14]. Complementing these automation tools, Rinaldi and Thun, in an infection control use case, develop handcrafted mappings from openEHR microbiology archetypes to FHIR resources and from FHIR to OMOP tables, performing an internal map-quality assessment to evaluate them, and documenting coverage gaps and context-dependent decisions [13].

These approaches demonstrate the feasibility and value of reusable, standards-based mappings and analytics-ready views over heterogeneous data, but they operate directly at the level of implementation artifacts and treat mappings largely as syntactic or structural transformation rules between schemas. The underlying conceptualizations and meaning-level relations remain implicit, leaving semantic validity dependent on case-by-case decisions and creating a risk of false agreement. By contrast, a shared, foundationally grounded reference model helps ensure that mapped data elements remain semantically preserved and comparable across standards.

Whereas the above work primarily operationalizes point-to-point interoperability through mappings between implementation artifacts (schemas, archetypes, and transformation rules), complementary work has proposed centralized or concept-oriented layers that reduce dependence on any single data model and make concept definitions reusable across representations. Gaudet-Blavignac et al. outline the Swiss Personalized Health Network's strategy, in which concept definitions serve as a central axis of interoperability—encoded in a "semantic framework" and represented and exchanged via RDF—while remaining deliberately model-independent [27]. Similarly, Bönisch et al. define a metadata crosswalk between FHIR, OMOP, Clinical Data Interchange Standards Consortium (CDISC), and openEHR based on reference metadata items and a priority scheme, which together function as a pragmatic intermediary layer for aligning heterogeneous metadata in FAIR clinical data integration workflows [28]. Zhang et al. build on the ISO/IEC 11179 metadata registry model to distinguish data element concepts from their representations. They operationalize this separation using Simple Standard for Sharing Ontological Mappings (SSSOM)-, RDF-, and Simple Knowledge Organization System (SKOS)-based mappings across multiple health data standards in a sex and gender use case, enabling queries at different levels of conceptual granularity [29]. At a more general Semantic Web integration layer, Dumontier et al. introduce the Semanticscience Integrated Ontology (SIO), a simple upper-level ontology and set of design patterns used to structure and integrate biomedical linked data and semantic web services [30]. Building on SIO, Alarcón-Moreno and Wilkinson's Clinical and Registry Entries Semantic Model (CARE-SM) proposes a Semantic Web representation for clinical and registry entries with a standardized and homogenized core structure, including a contextual metadata layer, to facilitate federated querying and reasoning and to support downstream mappings to other data models [31].

These contributions share with our work the premise that concepts and mappings should be treated as first-class entities and that interoperability mechanisms should separate meaning from representational choices; however, they typically prioritize pragmatic, model- or metadata-centric harmonization and do not aim to make foundational ontological commitments explicit. In contrast, the Health-RI Semantic Interoperability Initiative introduces the Health-RI Ontology (HRIO) as a UFO-grounded domain ontology that serves as a common semantic reference model and the Health-RI Mapping Vocabulary (HRIV) as an intentional mapping vocabulary. This combination is intended to turn standard-specific schema- and metadata-level mappings into semantically traceable alignments to a shared conceptualization, with the expected benefit of reducing false agreement and supporting systematic reuse across datasets, standards, and institutions.

High-level analyses and national strategies further situate this work within the broader policy and governance landscape. In the context of the European Health Data Space (EHDS), Palojoki and Vuokko review a range of semantic interoperability approaches and identify "semantic development goals" related to harmonization of heterogeneous standards, cross-system access, and meaningful secondary use of Electronic Health Record (EHR) data [32].

The Health-RI Semantic Interoperability Initiative can be viewed as a concrete, model-driven response to these goals and experiences: it retains the non-intrusive, standards-reusing character of such initiatives but adds explicit ontological commitments and meaning-level mapping relations. By grounding HRIO in UFO and using HRIV to type mappings and make representation relations explicit, our framework aims to enable semantic traceability and to anchor schema- and metadata-level mappings to a shared conceptualization. In doing so, it is intended to provide a semantic hub that can complement transformation-focused frameworks, concept- and metadata-oriented approaches, and policy-driven strategies, while addressing the semantic gaps identified above.

3. The Health-RI Semantic Interoperability Initiative

The Health-RI Semantic Interoperability Initiative tackles the semantic interoperability problem by making the intended meaning of concepts explicit and traceable across all the artifacts involved in data exchange. The Initiative's resources are openly available via its website (https://w3id.org/health-ri/semantic-interoperability) and its Git repository (https://w3id.org/health-ri/semantic-interoperability/git). Rather than assuming that two external implementation artifacts (e.g., schemas, taxonomies, ontologies, or vocabularies) agree because they use similar labels, textual definitions, codes, or even OWL axioms, the Initiative requires that every mapped concept in any such artifact be traced back to a shared, well-defined conceptualization captured in a common semantic reference model. In this sense, *semantic traceability* denotes the ability to follow and justify how the meaning of a concept is preserved across layers of representation—from phenomena, through conceptual models and computational ontologies, to external artifacts that are mapped to it.

3.1. Conceptual Foundations: Semantic Traceability

Conceptually, semantic traceability builds on two complementary theoretical pillars. First, the classic semiotic triangle (often referred to as the "triangle of reference") [33] distinguishes three elements. These are *(i)* a *referent* (in a domain of discourse), *(ii)* a conceptual *meaning* (i.e., a thought or reference), and *(iii)* a symbolic *expression* (i.e., a symbolic representation such as terms, codes, schema elements, and OWL classes) used to represent that meaning. In this paper, we treat the relevant conceptual meaning as a shared conceptualization made available as a common reference point to which different artifacts are explicitly connected, so that intended meanings can be compared and traced across representations. This conceptual framing, represented in Figure 1's left box, provides the basic structure along which traceability must be maintained, from phenomena, via shared concepts, to their encoding in schemas and ontologies.

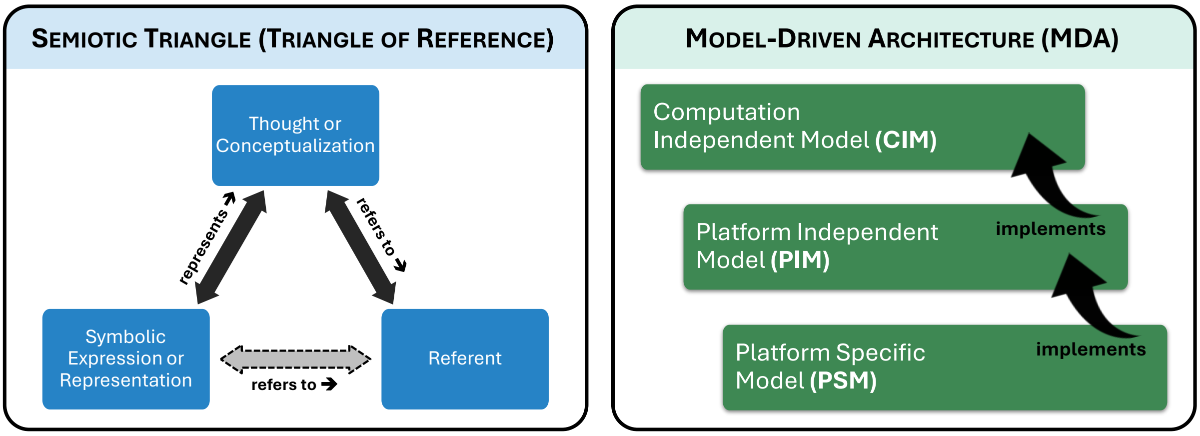

Figure 1 - Theoretical background for semantic traceability: the semiotic triangle (left) distinguishes meaning from representation, and Model-Driven Architecture (right) structures how system specifications are related across the Computation Independent Model (CIM), Platform Independent Model (PIM), and Platform Specific Model (PSM).

Here, we use the term "concept" to denote a domain-level unit of meaning in a shared conceptualization (e.g., Person, Biological Sample), distinct from both the phenomena it refers to (the referent) and the symbolic or data-level structures (e.g., codes, labels, data elements) that represent it. In ontology engineering, this triangle has been explicitly adopted at both conceptual and implementation levels—for instance, in Guizzardi's use of it to structure ontologically well-founded conceptual models [20] and in Gangemi's semiotics ontology design pattern for Semantic Web representations [34].

Second, the Model-Driven Architecture (MDA) approach [35], as shown in Figure 1 (right), structures system descriptions into a Computation Independent Model (CIM), a Platform Independent Model (PIM), and a Platform Specific Model (PSM). Figure 1 illustrates the refinement relation in which an artifact at a lower layer provides a concrete realization of the specifications captured at the layer above, through an implementation relationship between the layers.

In our setting, the CIM corresponds to the shared domain conceptualization made explicit in HRIO, whereas multiple PIM artifacts may coexist: (i) HRIO's computational OWL implementation (as the PIM that implements the CIM), and (ii) standard-level models such as FHIR, OMOP, and openEHR (as PIMs for their respective ecosystems). Concrete deployments—such as local database schemas, ETL code, message implementations, and runtime services—constitute PSMs. In combination with Semantic Web technologies, MDA has been proposed as a promising way to resolve semantic interoperability issues in the health data domain [3]. This clarifies how domain conceptualizations are progressively refined into technical artifacts, with requirements at the CIM level remaining traceable to PIM and PSM constructs across layers [35]. The separation of concerns supports systematic refinement and reuse, improves traceability between conceptual and technical artifacts, and facilitates the controlled evolution of implementations [3], [35].

Within this semiotic and model-driven perspective, the Initiative's architecture enables semantic traceability through two tightly related conceptual artifacts. First, a common semantic reference model captures shared domain meaning and operationalizes it in a machine-processable form, pairing an expressive conceptual specification with a computationally tractable OWL implementation. Second, a mapping vocabulary expresses how external concepts relate to this shared meaning. The following subsections outline these artifacts and explain how they jointly sustain semantic traceability.

3.2. Common Semantic Reference Model: The Health-RI Ontology

At the core of the architecture is the Health-RI Ontology (HRIO), which serves as a common semantic reference model for integrating heterogeneous health and life sciences data. In our architecture, HRIO is a *domain reference ontology* positioned above implementation standards such as FHIR, OMOP, and openEHR as a semantic layer, while itself being ontologically grounded in the foundational ontology UFO. Published at the persistent namespace IRI https://w3id.org/health-ri/ontology# (prefix "*hrio:*"), HRIO provides the shared reference point to which standards and schemas map their concepts, so that intended meanings can be made clear and explicit for interoperability across systems.

HRIO is realized as two tightly related artifacts with distinct but complementary roles. At the conceptual level, an OntoUML ontology, hereafter called HRIO OntoUML, is used as the Computation Independent Model. OntoUML is a conceptual modeling language grounded in the Unified Foundational Ontology (UFO) [20], [23] and is employed to construct a semantically rich, ontologically well-founded common reference model that serves as a semantic anchor for participating schemas [18].

Rather than treating semantic alignment as a terminological exercise, HRIO treats it as an explicit accounting of ontological commitments. In practice, shared labels, codes, and textual definitions may mask divergent conceptualizations [18], [22]. Moreover, commonly used representation formalisms (including ontology languages) are often compatible with multiple ontological readings, so the intended commitments are frequently underdetermined by surface form and cannot be reliably inferred from labels, definitions, or axioms alone [18].

By adopting UFO's well-defined ontological distinctions and constraints, OntoUML enables the explicit "unpacking" of real-world semantics and makes modeling commitments discussable, reviewable, and comparable across artifacts [18], [22], [23]. This model supports domain understanding, problem-solving, and meaning negotiation, thereby facilitating informed design decisions and communication with domain experts [22], [23]. In addition, UFO-based conceptual modeling is explicitly intended to support ontological analysis and conceptual clarification—i.e., precisely the activities needed when interoperability cannot rely on terminology alone [23].

At the computational level, HRIO is realized as HRIO gUFO/OWL, a gUFO-based OWL ontology that serves as the Platform Independent Model. gUFO ("gentle UFO") provides a lightweight OWL implementation of UFO [25]. Accordingly, HRIO gUFO/OWL is generated by exporting the HRIO OntoUML reference model through the transformation provided by the *ontouml-vp-plugin* tool (https://w3id.org/ontouml/vp-plugin). The resulting ontology provides a machine-processable implementation that aims to preserve the intended conceptual semantics specified in OntoUML, within the constraints of OWL.

In the Initiative's architecture, every class in HRIO gUFO/OWL is understood as implementing one of the HRIO OntoUML concepts. In our setting, unlike the general case in which OWL-level representations may admit multiple ontological readings (especially when their intended ontological commitments are not made explicit), the intended interpretation of HRIO gUFO/OWL is established by the CIM-level specification and maintained through the controlled CIM to PIM transformation (OntoUML to gUFO). This is possible because the transformation targets gUFO, whose axiomatization reflects the UFO commitments already captured in the OntoUML model. Thus, OWL classes are not meant to introduce new meanings; they function as computational carriers of the meanings established in HRIO OntoUML and provide stable, machine-processable targets to which external artifacts can be linked for query and interoperability purposes.

Figure 2 provides an abstract view of semantic traceability in this architecture. On the left (semiotic view), the globe denotes referents (e.g., human beings taken generically, as a class rather than as individuals). The red thought cloud denotes a shared conceptualization of those referents—a domain-level interpretation that mediates between reality and its representations. "Shared" here does not mean that all agents have identical mental models; rather, it denotes a conceptualization made available as a common reference point to which different artifacts are explicitly connected, so that they refer to the same underlying interpretation. The semiotic view also shows the basic relations involved: the OntoUML model (as a representation) represents the conceptualization and (indirectly, via the conceptualization) refers to the referent, and the conceptualization itself (directly) refers to the referent.

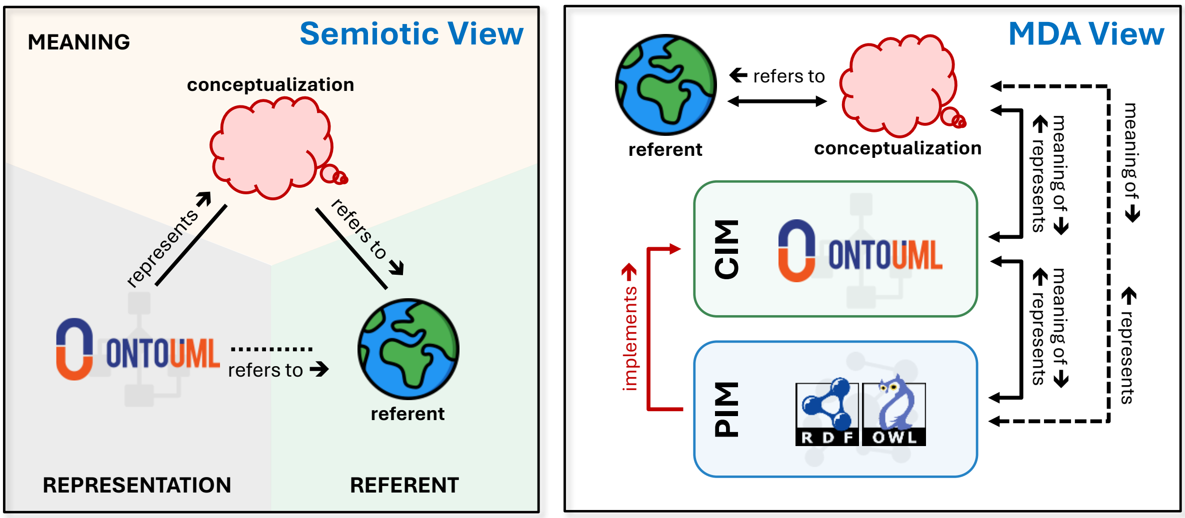

Figure 2 - Semantic traceability in the Health-RI Semantic Interoperability Initiative's architecture. The semiotic view (left) shows OntoUML as a representation of a shared conceptualization of the referent domain. The MDA view (right) shows HRIO OntoUML as the CIM and the HRIO gUFO/OWL ontology as the PIM, with the PIM implementing the CIM.

On the right (MDA view), the shared conceptualization is made explicit in an OntoUML reference model at the CIM layer (HRIO OntoUML, in the upper box) and is implemented by a gUFO-based OWL ontology at the PIM layer (HRIO gUFO/OWL, in the lower box). The red arrow labeled "implements" captures the (computational) MDA relation between these artifacts. The black connectors summarize non-computational semantic relations inherited from the semiotic view: they show how the CIM and the PIM are interpreted as representing the shared conceptualization. Solid black connectors indicate direct semantic relations, whereas dashed black connectors indicate indirect ones.

3.3. The Health-RI Mapping Vocabulary

The Health-RI Mapping Vocabulary (HRIV), published at the persistent namespace IRI https://w3id.org/health-ri/mapping-vocabulary# (prefix "*hriv:*"), provides the instrument for expressing how external concepts relate to the meanings defined in HRIO. Although SKOS provides widely used mapping properties, they are designed to relate SKOS concepts and characterize "match" in terms of practical interchangeability (not as formal meaning-level commitments). In our setting, we require mapping properties to have a precise and explicit intentional (definitional) interpretation, applicable to heterogeneous semiotic expressions (e.g., OWL classes, schema elements, codes) mapped to HRIO meanings. For example, the SKOS Reference [36] defines *skos:exactMatch* as a property that should be used to indicate "*a high degree of confidence that the concepts can be used interchangeably across a wide range of information retrieval applications*", which is less explicit than the meaning-level commitments we require for semantic traceability.

Conceptually, it grounds mappings in an intentional, semiotic interpretation, and computationally it is implemented as an OWL vocabulary defining a family of mapping properties and supporting the set-theoretic reasoning used with OWL and SKOS. In other words, HRIV is designed to be meaning-centric while remaining compatible with practical SKOS-based mapping practices. In the Initiative's architecture, external standards, schemas, and ontologies contribute the source expressions being mapped, while HRIO concepts play the role of the target meanings (with inverse properties available for querying in the opposite direction).

HRIV introduces an abstract (i.e., not intended for direct use) superproperty *hriv:meaningMappingRelation* as a general semantic mapping relation from a semiotic expression (e.g., a class in an external ontology) to an HRIO concept (operationalized in HRIO gUFO/OWL as an OWL class). The property hierarchy is deliberately closed and controlled as a conformance requirement, rather than an OWL-enforced logical constraint: HRIV mappings are intended to use only HRIV-defined subproperties of *hriv:meaningMappingRelation*, so that each mapping carries an explicit, curated mapping type and remains consistently interpretable.

For semantic traceability, the core object properties can be summarized as follows. *hriv:hasExactMeaning* is a specialized subproperty of *semiotics:expresses* [34], indicating that an expression's intended semantics are fully and precisely defined by a specific linked HRIO meaning. Exact meaning mappings are preferred when appropriate. *hriv:hasBroaderMeaningThan* and *hriv:hasNarrowerMeaningThan* are used when an exact-meaning alignment is not justified (or not available), to make the direction of semantic mismatch explicit and thereby reduce the space of potentially invalid interpretations.

hriv:hasBroaderMeaningThan indicates that an expression is broader in scope than the HRIO target it is linked to. It is used when the expression includes the linked HRIO concept but is not limited to it (e.g., "Adult patient" mapped to "Pregnant adult patient"), so the linked HRIO concept is a narrower (best-available) target rather than an exact match. Conversely, *hriv:hasNarrowerMeaningThan* indicates that an expression is narrower in scope than the HRIO target it is linked to. It is used when the expression is limited to a subset of what the linked HRIO concept covers (e.g., "Left femur fracture" mapped to "Femur fracture"), so the linked HRIO concept is a broader (best-available) target rather than an exact match. In both cases, the chosen mapping type makes the direction of non-equivalence explicit, helping data integrators avoid treating the mapped concepts as interchangeable.

Derived cross-standard relationships reflect the asserted mapping type (i.e., exact-meaning alignment versus broader/narrower-meaning alignment) and should be interpreted accordingly, rather than as equivalence, unless an exact-meaning alignment applies. Each of these mapping relations is aligned via *rdfs:subPropertyOf* to a corresponding SKOS mapping relation, so that Health-RI mappings can participate in standard SKOS-based tooling and workflows. However, the HRIV predicates remain the authoritative carriers of the intended (definitional) semantics: any entailed SKOS mapping assertions should be understood as a derived "SKOS view" for interoperability and discovery, and SKOS-level entailments (e.g., those enabled by *skos:exactMatch* being transitive) must not be used to change the intended interpretation of an HRIV mapping.

Specifically, HRIV's primary commitment is intentional: mappings relate expressions to meanings in a semiotic, definitional sense. In the OWL implementation, mapping assertions are expressed as object-property assertions between individuals denoting expressions and meanings. When the IRIs of OWL classes are used in these assertions, they are intended to be interpreted in OWL 2 DL via punning. Therefore, HRIV mappings should not be interpreted as OWL class axioms: they support traceable, machine-processable meaning links and SKOS-style workflows, but they do not imply class equivalence (*owl:equivalentClass*), identity (*owl:sameAs*), or subsumption (*rdfs:subClassOf*) entailments between the mapped classes.

Because every class in the HRIO gUFO/OWL ontology is semantically defined by its counterpart in the HRIO OntoUML reference model, HRIV mappings make explicit the meaning-level link from external artifacts to HRIO meanings, represented by HRIO gUFO/OWL classes. When a concept in an external ontology such as *std1:Human* is mapped to *hrio:Person* using *hriv:hasExactMeaning*, it is intended to be interpreted as embodying the same conceptualization originally defined in the HRIO OntoUML model.

By grounding both the HRIO OntoUML and the HRIO gUFO/OWL ontologies in the same meaning and by using HRIV to link external concepts to that meaning, the risk of the OWL ontology drifting from the original intent can be mitigated. This supports semantic traceability in both directions: from concepts in external artifacts (via HRIV mappings) to HRIO gUFO/OWL and the corresponding meanings in HRIO OntoUML (and ultimately the referent domain), and from HRIO meanings back to their mapped external representations.

In practice, HRIV supports two complementary mapping approaches. When external standards, schemas, or ontologies are outside Health-RI's editorial control (for example, national standards, or public vocabularies), mappings are typically maintained non-invasively in a separate mapping artifact (e.g., in a SSSOM Mapping Set), since Health-RI cannot modify those sources. When an external artifact is under the editorial control of its authors or maintainers, HRIV mapping assertions may be embedded directly in the artifact itself (as ontology annotations or axioms), creating tighter integration while still relying on the same meaning-centric semantics. This supports non-invasive and flexible adoption: contributors can maintain mappings in separate mapping sets and embed mappings only when editorial control and governance permit.

4. Simple Example of Semantic Traceability

To illustrate how the architectural components introduced in Section 3 support semantic traceability in practice, Figure 3 presents a simplified example centered on the concept Person and its sex- and gender-related specializations. The HRIO OntoUML and HRIO gUFO/OWL elements shown in this figure constitute a small, simplified excerpt of Health-RI's conceptualization for human beings, specifically concerning sex and gender, as captured in these ontologies.

Figure 3 is organized into three vertical regions: HRIO OntoUML on the left, the HRIO gUFO/OWL ontology in the middle, and external ontologies on the right. Both HRIO OntoUML and HRIO gUFO/OWL are on a yellow background to indicate Health-RI ownership, and the external ontologies are on a gray background indicating non-Health-RI ownership. While the OntoUML model is inside a blue round rectangle representing that it is a conceptual model, both HRIO gUFO/OWL and external ontologies are in a green round rectangle indicating they are computational artifacts.

In HRIO OntoUML, the model contains the class *Person* (stereotyped as a UFO Kind) and some of its specializations, such as *Male Person*, *Karyotypical Male*, and *Person with Male Gender*. In HRIO gUFO/OWL, the ontology contains corresponding OWL classes *hrio:Person*, *hrio:MalePerson*, *hrio:KaryotypicalMale*, and *hrio:PersonWithMaleGender*, related by *rdfs:subClassOf* axioms that mirror the conceptual hierarchy. The blue dashed arrow labeled "(directly) represents" indicates that *hrio:Person* is assumed to directly represent the same meaning as *Person*. The dashed style indicates that this is a non-computational relation: it is not expressed in OWL, but rather a semantic assumption stating that the OWL class implements the OntoUML concept.

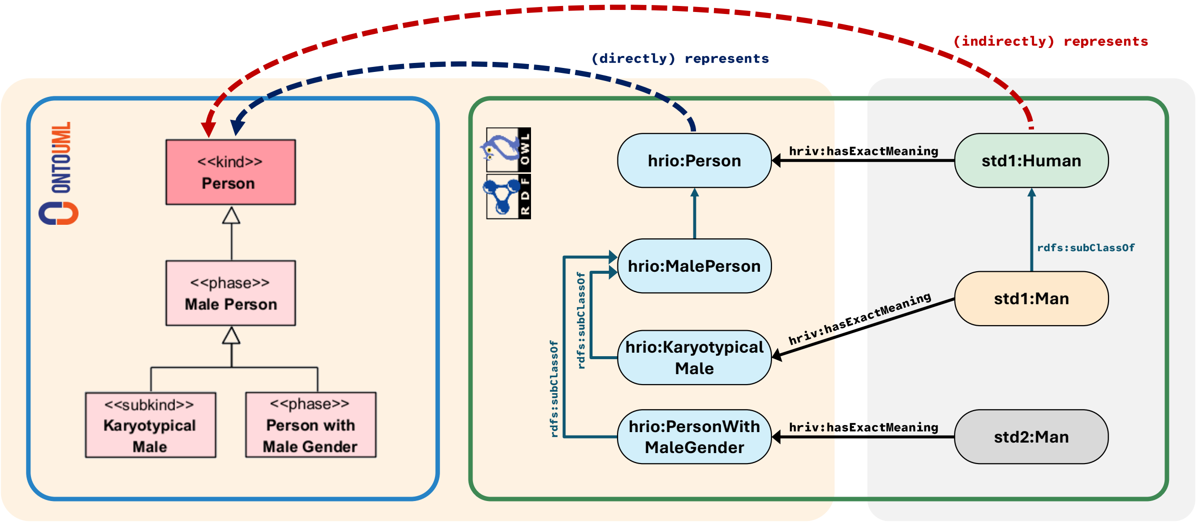

Figure 3 - Example of semantic traceability: a fragment of HRIO OntoUML for Person's sex and gender is implemented in HRIO gUFO/OWL, and external ontologies' OWL classes are aligned via HRIV mappings, distinguishing direct and indirect representation relations. The semantics of the external classes are traceable to their corresponding meanings modeled in OntoUML.

Two illustrative external ontologies are used to facilitate the explanation of semantic traceability as enabled by the framework. One standard (prefix "*std1:*") distinguishes between *std1:Human* and the more specific *std1:Man* (a karyotype-based reading). Another standard (prefix "*std2:*") contains a single class, also labeled "Man" (*std2:Man)*, which is defined as a gender-based reading of "man". These external artifacts are outside Health-RI's editorial control but participate in semantic interoperability via HRIV mappings.

In the OWL ontologies in Figure 3, black solid arrows depict HRIV mapping relations, while blue solid arrows depict *rdfs:subClassOf* axioms. For instance, *std1:Human* is linked to *hrio:Person* via *hriv:hasExactMeaning*, expressing that the standard's notion of human beings is taken to match exactly the shared conceptualization captured in the OntoUML *Person* and implemented as *hrio:Person*. *std1:Man* is linked to *hrio:KaryotypicalMale* using *hriv:hasExactMeaning*, reflecting that this external concept corresponds to a specific biological (karyotype-based) reading of male persons. By contrast, *std2:Man* is mapped to *hrio:PersonWithMaleGender* via *hriv:hasExactMeaning*, making explicit that this external class is intended to capture the gender-based meaning represented in HRIO (despite its potentially ambiguous label).

Once both external classes are linked to HRIO meanings, their semantic relationship can be assessed by inspecting (and, computationally, querying and reasoning over) the HRIO OntoUML model and its corresponding HRIO gUFO/OWL implementation. Here, the asserted mapping type makes explicit how each external class is intended to relate to the shared HRIO meaning and, consequently, how any derived cross-standard relationship should be interpreted, as a mapping-level assessment grounded in HRIO (rather than as an OWL DL class-level entailment between the external classes). The assessment is grounded in HRIO's class hierarchy, relations, and associated constraints (e.g., disjointness), rather than relying on labels or textual descriptions alone. For example (not shown in the fragment in Figure 3), the model can make explicit that a karyotypical female person may still be a person with male gender (via self-assigned gender), whereas the karyotype-based reading does not account for this possibility. Consequently, despite sharing the same label, *std1:Man* and *std2:Man* should not be treated as directly interoperable, and these differences must be considered when designing data exchange between systems that implement them.

Finally, the red dashed arrow labeled "(indirectly) represents" shows how external ontology classes should be understood as indirectly representing OntoUML concepts via their HRIV links to

HRIO meanings. This indirect representation is not encoded as a computational relation in OWL; rather, it is derived from the combination of implementation and mapping assumptions. Together, the blue and red dashed arrows and the HRIV mappings illustrate how semantic traceability is maintained from the shared conceptualization represented in the HRIO OntoUML reference model, through the HRIO gUFO/OWL ontology, to heterogeneous external standards and schemas. For readability, these two dashed relations are shown only once in Figure 3, but they apply to all mapped concepts in the figure (e.g., the HRIO OntoUML class *Person with Male Gender* is directly represented by *hrio:PersonWithMaleGender* and indirectly represented by *std2:Man*).

5. Discussion and Future Work

The Health-RI Semantic Interoperability Initiative proposes a model-driven, ontology-based framework that is intended to enable semantic traceability across conceptual and computational layers to mitigate semantic misalignment (*false agreement*), thereby supporting interoperable reuse of health data. It combines an OntoUML common reference model (HRIO OntoUML) with a gUFO-based OWL implementation (HRIO gUFO/OWL) and an intentional (definitional) mapping vocabulary (HRIV). The framework is intended to shift effort from repeatedly creating and maintaining pairwise mappings toward reusable, meaning-explicit alignments by providing a single semantic hub to which standards and local schemas can be aligned.

The *Person* example illustrates how this hub can clarify the distinct conceptualizations embedded in existing standards and support preserving semantic intent as meanings move from shared conceptualizations to OWL ontologies and external implementation artifacts. The artifacts produced by the Initiative are openly published with persistent identifiers, reinforcing their role as reusable community assets rather than project-specific artifacts. Current deliverables include the HRIO OntoUML reference model, the HRIO gUFO/OWL implementation, the HRIV mapping vocabulary, and a worked example. Documentation of the Initiative's approach and of the published artifacts is available via the project website and its Git repository.

While these results illustrate the intended benefits of the framework, this manuscript does not present an empirical evaluation of mapping-effort reduction, error reduction, or user comprehension gains. Such an evaluation is part of the Initiative's future work.

Regarding limitations and adoption considerations, adopting the framework here proposed may increase the initial effort for contributors in the short term, since new standards and local schemas still require curated meaning mappings. The expected benefit is reduced duplication of mapping work across repeated integration settings, as later efforts can reuse existing HRIO-aligned mappings. Nevertheless, mappings are maintained per source artifact and release and must be reviewed when source or HRIO definitions evolve. The long-term success of the approach therefore depends on sustained community engagement and governance to curate and maintain mappings as artifacts evolve.

Beyond this evaluation, future work focuses on consolidating and expanding this architecture through concrete use cases in the health and life sciences domain. On the content side, HRIO will be incrementally extended by selecting prioritized concepts and taking each of them through a structured cycle of modeling, formalization, review, and publication, with continued involvement of domain experts to refine ontological commitments and resolve terminology and scope decisions.

On the technical side, we plan to strengthen the HRIV mapping lifecycle around those concepts—authoring, curation, review, validation, and visualization—supported by reusable templates and increasing automation, and to improve packaging and publishing of ontology and mapping releases with consistent metadata and documentation. Concurrently, we will expand user-facing support assets (e.g., query examples and high-level metrics/monitoring views) and integrate the artifacts into data integration pipelines, query services, and FAIR data publishing workflows.

Finally, responding to the identified lack of empirical evidence on the advantages and drawbacks of foundational ontologies in biomedical research [37], the approach could be evaluated in real-world integration scenarios. Such an evaluation could assess its impact on mapping effort, error reduction, and user comprehension and could also inform governance for the sustainable evolution of HRIO and its mappings. Concretely, a study could be designed to compare HRIO-aligned meaning-mapping to direct pairwise mapping practices (e.g., SSSOM mapping sets or SKOS-based mapping assertions) in tasks such as initial mapping creation, mapping revision after new releases of standards and local schemas, and cross-standard query formulation. Measures of interest could include time-on-task, the number of mapping revisions and reviewer disagreements, and the frequency of expert-identified false-agreement cases, complemented by short comprehension questionnaires.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT 5.2 for grammar and spell checking. After using this tool, the authors reviewed and edited the content as needed, and they take full responsibility for the publication's content.

References

[1] M. D. Wilkinson *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Sci. Data*, vol. 3, no. 1, p. 160018, Mar. 2016, doi: 10.1038/sdata.2016.18.

[2] J. Piera-Jiménez, H. Leslie, R. Dunscombe, and C. Pontes, "Interoperability in the Context of Integrated Care," in *Handbook of Integrated Care*, V. Amelung, V. Stein, E. Suter, N. Goodwin, R. Balicer, and A.-S. Beese, Eds., Cham: Springer Nature Switzerland, 2024, pp. 1–22. doi: 10.1007/978-3-031-25376-8_101-1.

[3] R. Yadav, S. Murria, and A. Sharma, "A research review on semantic interoperability issues in electronic health record systems in medical healthcare," in *IoT-Based Data Analytics for the Healthcare Industry*, Elsevier, 2021, pp. 123–138. doi: 10.1016/B978-0-12-821472-5.00009-0.

[4] B. H. De Mello *et al.*, "Semantic interoperability in health records standards: a systematic literature review," *Health Technol.*, vol. 12, no. 2, pp. 255–272, Mar. 2022, doi: 10.1007/s12553-022-00639-w.

[5] N. Capuano, P. Foggia, L. Greco, and P. Ritrovato, "A Linked Data Application for Harmonizing Heterogeneous Biomedical Information," *Appl. Sci.*, vol. 12, no. 18, p. 9317, Sep. 2022, doi: 10.3390/app12189317.

[6] I. Bhushan, K. Anandampillai, and S. Agarwal, "Navigating complexities: agile digital health initiatives in developing countries," *Oxf. Open Digit. Health*, vol. 2, p. oqae032, Jan. 2024, doi: 10.1093/oodh/oqae032.

[7] K. Ndlovu, T. Mogotlhwane, M. Mars, and R. E. Scott, "E-Health Interoperability Landscape: Botswana," in *Environment and Water Resource Management / 837: Health Informatics / 838: Modelling and Simulation / 839: Power and Energy Systems*, Gaborone, Botswana: ACTAPRESS, 2016. doi: 10.2316/P.2016.837-010.

[8] HL7 International, "FHIR Release 5." 2023. Accessed: Dec. 01, 2025. [Online]. Available: https://www.hl7.org/fhir/R5/

[9] openEHR Foundation, "Architecture Overview." 2020. Accessed: Dec. 01, 2025. [Online]. Available: https://specifications.openehr.org/releases/BASE/latest/architecture_overview.html

[10] SNOMED International, "SNOMED CT Starter Guide." 2025. Accessed: Dec. 01, 2025. [Online]. Available: https://docs.snomed.org/snomed-ct-practical-guides/snomed-ct-starter-guide/

[11] Observational Health Data Sciences and Informatics (OHDSI), "OMOP Common Data Model." 2014. Accessed: Dec. 01, 2025. [Online]. Available: https://ohdsi.github.io/CommonDataModel/

[12] S. Kohler *et al.*, "Eos and OMOCL: Towards a seamless integration of openEHR records into the OMOP Common Data Model," *J. Biomed. Inform.*, vol. 144, p. 104437, Aug. 2023, doi: 10.1016/j.jbi.2023.104437.

[13] E. Rinaldi and S. Thun, "Mapping from openEHR to FHIR and OMOP CDM to support interoperability for infection control," *GMS Med. Inform.*, p. Biometrie und Epidemiologie; 17(2):Doc07, Apr. 2021, doi: 10.3205/MIBE000221.

[14] S. Kohler *et al.*, "FHIRconnect: Towards a seamless integration of openEHR and FHIR," Nov. 18, 2025, *arXiv*: 2511.14618. doi: 10.48550/arXiv.2511.14618.

[15] E. Rinaldi and S. Thun, "From OpenEHR to FHIR and OMOP Data Model for Microbiology Findings," in *Studies in Health Technology and Informatics*, J. Mantas, L. Stoicu-Tivadar, C.

Chronaki, A. Hasman, P. Weber, P. Gallos, M. Crişan-Vida, E. Zoulias, and O. S. Chirila, Eds., IOS Press, 2021. doi: 10.3233/SHTI210189.

[16] HL7 Vulcan, "FHIR to OMOP." 2025. Accessed: Dec. 15, 2025. [Online]. Available: https://hl7vulcan.org/projects/fhir-to-omop/

[17] R. K. Mithiri, "A Semantic Interoperability Framework for Cross-Institutional Integration of Heterogeneous Electronic Health Records," *Int. J. Eng. Technol. Res. Dev.*, vol. 6, no. 3, pp. 25–31, May 2025.

[18] G. Guizzardi, "Ontology, Ontologies and the 'I' of FAIR," *Data Intell.*, vol. 2, no. 1–2, pp. 181–191, Jan. 2020, doi: 10.1162/dint_a_00040.

[19] N. Guarino, *Formal Ontology in Information Systems: Proceedings of the 1st International Conference June 6-8, 1998, Trento, Italy*, 1st ed. NLD: IOS Press, 1998.

[20] G. Guizzardi, "Ontological foundations for structural conceptual models," PhD Thesis - Research UT, graduation UT, University of Twente, 2005.

[21] F. J. Ekaputra, M. Sabou, E. Serral, E. Kiesling, and S. Biffl, "Ontology-Based Data Integration in Multi-Disciplinary Engineering Environments: A Review," *Open J. Inf. Syst. OJIS*, vol. 4, no. 1, pp. 1–26, 2017.

[22] G. Guizzardi and N. Guarino, "Explanation, semantics, and ontology," *Data Knowl. Eng.*, vol. 153, p. 102325, Sep. 2024, doi: 10.1016/j.datak.2024.102325.

[23] G. Guizzardi, A. Botti Benevides, C. M. Fonseca, D. Porello, J. P. A. Almeida, and T. Prince Sales, "UFO: Unified Foundational Ontology," *Appl. Ontol.*, vol. 17, no. 1, pp. 167–210, Mar. 2022, doi: 10.3233/AO-210256.

[24] M. A. Saberi, H. Mcheick, and M. Adda, "From Data Silos to Health Records Without Borders: A Systematic Survey on Patient-Centered Data Interoperability," *Information*, vol. 16, no. 2, p. 106, Feb. 2025, doi: 10.3390/info16020106.

[25] J. P. A. Almeida, G. Guizzardi, T. P. Sales, and R. A. Falbo, "gUFO: A Lightweight Implementation of the Unified Foundational Ontology (UFO)." 2019. Accessed: Dec. 01, 2025. [Online]. Available: http://purl.org/nemo/doc/gufo

[26] G. Xiao *et al.*, "FHIR-Ontop-OMOP: Building clinical knowledge graphs in FHIR RDF with the OMOP Common data Model," *J. Biomed. Inform.*, vol. 134, p. 104201, Oct. 2022, doi: 10.1016/j.jbi.2022.104201.

[27] C. Gaudet-Blavignac, J. L. Raisaro, V. Touré, S. Österle, K. Crameri, and C. Lovis, "A National, Semantic-Driven, Three-Pillar Strategy to Enable Health Data Secondary Usage Interoperability for Research Within the Swiss Personalized Health Network: Methodological Study," *JMIR Med. Inform.*, vol. 9, no. 6, p. e27591, Jun. 2021, doi: 10.2196/27591.

[28] C. Bönisch, D. Kesztyüs, and T. Kesztyüs, "Harvesting metadata in clinical care: a crosswalk between FHIR, OMOP, CDISC and openEHR metadata," *Sci. Data*, vol. 9, no. 1, p. 659, Oct. 2022, doi: 10.1038/s41597-022-01792-7.

[29] S. Zhang, R. Cornet, and N. Benis, "Cross-Standard Health Data Harmonization using Semantics of Data Elements," *Sci. Data*, vol. 11, no. 1, p. 1407, Dec. 2024, doi: 10.1038/s41597-024-04168-1.

[30] M. Dumontier *et al.*, "The Semanticscience Integrated Ontology (SIO) for biomedical research and knowledge discovery," *J. Biomed. Semant.*, vol. 5, no. 1, p. 14, Dec. 2014, doi: 10.1186/2041-1480-5-14.

[31] P. Alarcón-Moreno and M. D. Wilkinson, "Take CARE of your patient data. Clinical And Registry Entries (CARE) Semantic Model," in *SWAT4HCLS 2024: Semantic Web Applications and Tools for Health Care and Life Sciences 2024*, M. Roos, A. Jacobsen, A. Splendiani, M. S. Marshall, A. Waagmeester, L. J. García Castro, K. Wolstencroft, K. Hettne, and R. Vos, Eds., in CEUR Workshop Proceedings, vol. 3890. Leiden, The Netherlands: CEUR-WS.org, 2024, pp. 116–123. Accessed: Dec. 15, 2025. [Online]. Available: https://ceur-ws.org/Vol-3890/paper-11.pdf

[32] S. Palojoki and R. Vuokko, "Semantic Interoperability of EHRs: Review of Approaches for Enhancing the Goals of European Health Data Space," Apr. 26, 2023, *Public Health and Healthcare*. doi: 10.20944/preprints202304.0957.v1.

[33] C. K. Ogden and I. A. Richards, *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*. London: Kegan Paul, Trench, Trubner & Co., 1923.

[34] A. Gangemi, "Semiotics Ontology Design Pattern." 2007. [Online]. Available: http://www.ontologydesignpatterns.org/cp/owl/semiotics.owl

[35] J. Miller and J. Mukerji, "MDA Guide Version 1.0.1," Object Management Group, omg/03-06-01, Jun. 2003. [Online]. Available: http://www.omg.org/cgi-bin/doc?omg/03-06-01.pdf

[36] A. Miles and S. Bechhofer, “SKOS Simple Knowledge Organization System Reference,” World Wide Web Consortium (W3C), W3C Recommendation REC-skos-reference-20090818, Aug. 2009. [Online]. Available: https://www.w3.org/TR/skos-reference/

[37] C. H. Bernabé *et al.*, “The use of foundational ontologies in biomedical research,” *J. Biomed. Semant.*, vol. 14, no. 1, p. 21, Dec. 2023, doi: 10.1186/s13326-023-00300-z.