



Semantic interoperability in health records standards: a systematic literature review

Blanda Helena de Mello¹ · Sandro José Rigo¹ · Cristiano André da Costa¹  · Rodrigo da Rosa Righi¹ · Bruna Donida¹ · Marta Rosecler Bez¹ · Luana Carina Schunke¹

Received: 19 October 2021 / Accepted: 7 January 2022 / Published online: 26 January 2022
© The Author(s) under exclusive licence to International Union for Physical and Engineering Sciences in Medicine (IUPESM) 2022

Abstract

The integration and exchange of information among health organizations and system providers are currently regarded as a challenge. Each organization usually has an internal ecosystem and a proprietary way to store electronic health records of the patient's history. Recent research explores the advantages of an integrated ecosystem by exchanging information between the different inpatient care actors. Many efforts seek quality in health care, economy, and sustainability in process management. Some examples are reducing medical errors, disease control and monitoring, individualized patient care, and avoiding duplicate and fragmented entries in the electronic medical record. Likewise, some studies showed technologies to achieve this goal effectively and efficiently, with the ability to interoperate data, allowing the interpretation and use of health information. To that end, semantic interoperability aims to share data among all the sectors in the organization, clinicians, nurses, lab, the entire hospital. Therefore, avoiding data silos and keep data regardless of vendors, to exchange the information across organizational boundaries. This study presents a comprehensive systematic literature review of semantic interoperability in electronic health records. We searched seven databases of articles published between 2010 to September 2020. We showed the most chosen scenarios, technologies, and tools employed to solve interoperability problems, and we propose a taxonomy around semantic interoperability in health records. Also, we presented the main approaches to solve the exchange problem of legacy and heterogeneous data across healthcare organizations.

Keywords Semantic interoperability · Health standard · Health record · EHR · Systematic review

1 Introduction

A semantically integrated health system allows sharing data among organizations and their internal ecosystem without missing the meaning. The search for semantic interoperability in health records and different clinical annotations is one of the main challenges for systems, being a constant objective of studies in the last years, such as [1–3]. Implementing interoperability can allow healthcare professionals to manage the complete electronic patient record, regardless of the organization that generated the clinical session entries [4]. As reinforced by Kim and Joshi [5], health record interoperability became a crucial issue in the healthcare scenario,

especially in the COVID-19 pandemic, to further disease control.

Semantic interoperability (SI) aims to share data among organizations or systems and ensure they understand and interpret data regardless of who is involved, using domain concepts, context knowledge, and formal data representation [6]. On the other hand, semantic interoperability also can be understood by taking a step back: semantics is the study of meaning, focused on the relationship between people and their words, which is essential to help people understand each other despite different experiences or viewpoints [7]. Interoperability is the ability of two or more systems to work together, regardless of its different interfaces, platforms, and technologies adopted [8].

The HIMSS [9]¹ – a global advisor supporting the transformation of the health ecosystem, with 480 provider organizations and more than 450 non-profit partners – had designed defined in three levels the interoperability technology: 1) Foundation;

✉ Cristiano André da Costa
cac@unisinis.br

¹ Software Innovation Laboratory, SOFTWARELAB
Universidade Do Vale Do Rio Dos Sinos -
Unisinis, São Leopoldo, Brazil

¹ <https://www.himss.org/who-we-are/> (HIMSS) Healthcare Information and Management Systems Society.

2) Structural and 3) Semantic. The Foundational level defined requirements to connect different systems and securely exchange data. The Structural level defines the format, syntax, and data to interpret at the field level; the Semantic level allows one to work with terminologies, vocabularies, standardized values publicly defined and makes a complete understanding meaning.

According to Gancel et al. [10], there are many challenges to overcome to achieve an electronic health record, such as understanding the wide variety of terms, addressing disambiguation, identifying and updating the concepts. That scenario is reinforced by [11, 12], where the authors had discussed the three fundamental levels of interoperability to improve the workflows across health information systems to allow true interoperability. To the healthcare area, interoperability means different systems, applications, and devices share, use, and process data from any place and keep the real meaning.

Using standards allows sharing data between clinicians, lab, hospital and, pharmacies regardless of vendor, achieving semantic interoperability. In other words, interoperability allows health information systems to work across organizational boundaries.

From the above, the main contributions of this paper are as follows:

- discussion about the state of the art of semantic interoperability in health records.
- introduce a taxonomy in semantic interoperability in electronic health records.
- recognize the main approaches commonly used to achieve semantic interoperability in EHR (Electronic Health Record) systems.
- expose the grown adoption using semantic web technologies combining international standards to solve semantic interoperability problems.

Thus, this systematic review aims to answer the following central question: what is the state of the art of semantic interoperability in health records in the sense of approaches and international health standards?

We organized the systematic review as follows. Section 0 presents a background of semantic interoperability and electronic health records. Follow, Sect. 0 shows the protocol defined, describing the inclusion/exclusion criteria and the quality assessment. Section 0 presents the results of the conducted review and the research questions answered, and Sect. 0 discusses some open questions and the main approaches used by the selected studies. Lastly, we discuss future directions to work in Sect. 0.

2 Background

According to ISO/TS18308 [13], an EHR aims to integrate health records in a processable format, securely storing and communicating. Using an information model

commonly accepted to exchange data, making it accessible to authorized users. That aims to ensure the patient's life-long integrated healthcare, efficiently with high-quality and security. The EHR structure holds the patient's health status, and that format must be processed digitally [14], maintaining patients' data throughout their life and stored accurately in a repository.

There are different data formats in EHR systems, such as structured data and non-structured textual data. EHR covers an extensive part of the medical histories, includes more patients' complete information and potential risk factors. In addition, it maintains patient health records and supports the provision of daily care in hospitals and primary care clinics [15]. Furthermore, it allows the reuse of patient data for many purposes, including managing individual patients, medical and health services research, and management of health care facilities [16].

Adopting an EHR is essential to manage healthcare and exchange data between healthcare organizations. An EHR allows communication among clinicians, nurses, laboratories, and hospitals despite different systems. Sharing data between health organizations and health agents must foster the correct interpretation, with the same precision and meaning adopted from the sender [17], achieving semantic interoperability. Therefore, semantic interoperability is the ability to share data between systems and ensure understanding at the concept's level of the domain [13].

Different health standards aim to enable data sharing among healthcare organizations. However, the adoption of standards still presents several challenges to achieving interoperability at the semantic level. The semantic web focuses on sharing data, integration, and reuse through ontologies, linked open data, and knowledge graphs to ensure the correct meaning of shared data. The semantic web is also known as one of the fundamental technologies to achieve semantic interoperability in health information systems [4], often using ontologies, a well-established technology to support knowledge-intensive tasks related to EHR systems [3].

3 Related work

Recently, the demand for systems that allow data interoperability at a semantic level has been an object of interest, mainly in health system providers. Different studies explore approaches to solving interoperability problems. However, there are difficulties in adopting health standards and tools to adequate data representation (ontologies, databases, clinical models) that ensure healthcare professionals efficiently manage the data.

The authors in [18] discuss interoperability in electronic health records from the management and business perspective.

In this way, they highlight how data integration and exchange across organizational boundaries can improve quality of care, work processes, and effectiveness to reduce costs and improve efficiency. Also, the authors had shown how related healthcare with other areas such as Telemedicine, Big Data, and Business Intelligence. On the technology side, it helps eliminate rework, reduce errors, and promote individualized patient care. On the citizens' side, it highlighted the support to create the primary public health initiatives, control, monitor diseases, reduce costs, and increase the effectiveness.

On the other hand, [19, 20] analyzed technological aspects regarding health integration, interoperability, and data exchange. The first work describes the advantages of implementing integrated data repositories, a clinical data warehouse that allows clinical research, specialized analysis, and advanced data processing. The second paper proposes a transnational model integrating health records. The study suggests adopting a widely used health standard called HL7 FHIR,² vocabularies, and terminologies.

The systematic review [21] discusses the strong trend in adopting standards. The authors showed interest in analyzing the literature according to ontologies, specifically fuzzy ontology (Fuzzy ontology). The study presents a comprehensive background to context for the leading health standards and their different structures, highlighting characteristics they may have in common. The article also named the adopted standards using the “e-Health Standard” term, whereas we keep the broad term as a health standard. Also, the authors had highlighted trends of semantic interoperability in four categories that contribute to identifying challenges and research opportunities: a) frameworks to solve SI problems, b) using ontologies to achieve interoperability issues, c) standards in an Interoperable EHR, d) barriers and the heterogeneous problem EHR semantic interoperability.

Furthermore, the studies developed by [20, 21] richly exposed the growth and interest of the healthcare sector in using standards for electronic medical records, breaking organizational barriers, and achieving interoperability among healthcare providers. Given this scenario, we conduct this review by scrutinizing the evolutions in adopting standards in recent years and the tools that eventually make up the environment for a semantically interoperable EHR.

Therefore, we consider it more coherent with the selected studies to observe the ecosystem involved in the construction and development of an EHR. Furthermore, understanding how secondary artifacts – semantic web tools, databases, terminologies – interact and impact project definition. Besides the standards and semantic web tools, we highlight ontologies not exclusively used to represent raw data but to represent clinical model structures.

² <https://www.hl7.org/fhir/>

4 Material and methods

This section describes the protocol used for the systematic literature review and shows the research questions designed to extract the information of interest from the selected studies. The research strategy for filtering and selecting the studies involved the adoption of inclusion and exclusion criteria. Finally, we show aspects of the quality assessment for the selected articles, later applied to answer the research questions.

4.1 Study design

The protocol followed the steps from previous works of [22, 23] and aimed to map relevant and recent research in semantic interoperability. A systematic literature review is a research method that allows the identification, evaluation, and interpretation of the studies without bias in the process.

- a) Research questions: introduce the research questions investigated.
- b) Search strategy: outline the strategy and libraries explored to collect data.
- c) Article selection: explain the criteria for selecting the studies.
- d) Quality assessment: describe the quality assessment applied to the selected studies.
- e) Data extraction: compare the selected studies and research questions.

4.2 Research questions

The research questions represent a fundamental part of the systematic review [22], as they allow directing the research and extracting information as needed, as they follow the topic of interest. Table 1 presents the research questions of interest in the systematic review. There are two sets of questions, the first global context questions, including an overview of all studies. The latter are specific questions to answer about each article, as shown below.

4.3 Search strategy

The search strategy aims to find studies to answer the research questions through the definition of keywords and the scope of the research. The correct selection of keywords ensures adequate research results on the databases. As guided by [22], we used the five PICOC questions (Population, Intervention, Comparison, Outcome, and Context) to define the scope of the research, as explained below.

Population involves keywords, related terms, and variants around the interest area. We used “semantic interoperability” as the central term in the search string shown in Frame 1. In

Table 1 Research questions to extract information of interest from the articles selected in this review

Group identifier	Issue
General questions (GQ)	
GQ1	What is the state of the art in health standards applied in health records?
GQ2	What are the challenge and open questions to semantic interoperability in health records?
Specific questions (SQ)	
SQ1	What are the health standards adopted in the studies?
SQ2	What are terminologies or health repositories used?
SQ3	What are the approaches used?
SQ4	What are the main security concerns used?
SQ5	What are the evaluation approaches used?

the Intervention, being “semantic interoperability” a general term, we used some other keywords to filter: health record, medical record, patient record, and hospital record, aiming to allow filtering semantic interoperability standards applied in health records. Comparison, on this research, aims to find different health standards allowing semantic interoperability in health records. The Outcomes determine relevant studies that can answer the research questions.

Also, we evaluated studies that explore the limitations of applying health standards. We define Context as the concern to identify the different application scenarios and data used on the proposal. This research focused on selecting studies concerned with proposing solutions to semantic interoperability problems in health records. In addition, we notice that different kinds of data come from the real world, and most come from a controlled environment (case study) or simulated data.

The research string formatted as shown in Frame 1 demonstrates the mandatory terms, such as semantic interoperability and his acronyms; health record and his acronyms; the term standard. At the selection step, the terms applied aim to filter the articles by title, abstract, and keywords.

Frame 1 This frame shows the search string defined to research.

“semantic interoperability” AND (“health record” OR “medical record” OR “patient record” OR “hospital record”) AND standard

4.4 Article selection

Next, the studies with neither a key term addressed nor a PICOC definition aligned, called impurities, are removed. Finally, qualify our results, and for that, we defined the exclusion criteria based on the research question:

- Exclusion criteria 1: the article is not in the range (search date 2010 to September 24, 2020).
- Exclusion criteria 2: the article does not address semantic interoperability or related acronyms (population criteria I).
- Exclusion criteria 3: the article does not address the health record, medical record, patient record, hospital record, and related acronyms (intervention criteria II).
- Exclusion criteria 4: the article does not address standard or related acronyms (comparison criteria III).
- Exclusion criteria 5: the article has less than six pages.
- Exclusion criteria 6: impurities articles (e.g., duplicate papers and non-English studies).

4.5 Quality assessment

The quality assessment consists of a filter focused on the quality and relevance of studies related to the interest area. The quality assessment ensures that selected articles have a relevant impact in the research area. As a criterion to filter the rest studies, we defined the h5-index score, equal to or higher than 28. If the article passes the cutoff, it can be accepted in this review..

4.6 Data extraction

After selecting articles to satisfy the inclusion and exclusion criteria and the quality assessment, the data extraction step occurs. The other articles represent the corpus qualified for full reading and analysis. The research questions target to extract information from selected studies according to the topic of interest, and data extraction aims to correlate the area of interest with the defined research questions. The studies propose different solutions to the research problem presented: *what health standards approaches can solve semantic interoperability problems*

Table 2 Follows we shows the quality assessment of article and related questions

Section	Description	Research Questions
Open Content	Title	GQ1, GQ2, SQ1
	Abstract	All questions
	Keywords	GQ1, SQ1
Article Content	Introduction	All questions
	Method	All questions
	Results	All questions
	Evaluation	All questions
	Conclusion	All questions

in health records? The questions have a limited response context to avoid bias, to respond objectively.

The methodology to answer the research questions proposes to extract the results as presented in Table 2, where the General Question (GQ) and Specific Question (SQ) have an expected location to search for data.

5 Results and discussion

The discussion of these different results extracted through the research questions defined, such as shown in Table 1, allows understanding the semantic interoperability scenario and the solutions usually applied in this scenario.

5.1 Conducting the search strategy

We chose seven different research bases to cover studies in the health and technology field, such as ACM Digital

Library, IEEE Xplore Library, Science Direct, Springer Link. Moreover, we added Google Scholar to cover studies outside those bases. Our criterion was the relevance of these databases concerning the health and information technology literature. The search step in the databases, as mentioned above, aimed to index the search for studies published in the last ten years. Each database presents a way of formatting the survey, which we respect and modify to suit, but we kept all the mandatory terms defined in the PICOC strategy.

Finally, after applying the queries to the search bases, we had 6,032 articles. The initial filter aims to remove patents and citations, non-English studies, which resulted in around 783, roughly because some patents also appeared as citations.

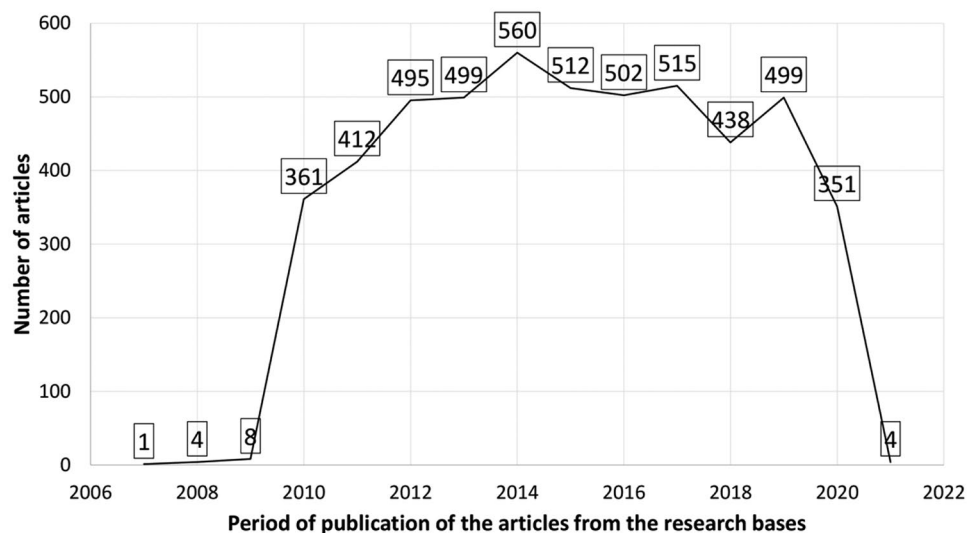
As shown in Fig. 1, by year, the published articles in this area have been an interest constant in recent years. For this systematic review, the cutoff was September 22nd, 2020.

5.2 Article selection

Figure 2 shows the selection steps, removing impurity studies unrelated to the area of interest. Usually, these impurities studies had references related to the research area or citations of related works, however, without directly informing the area of interest. In addition, articles with less than six pages and no abstract, about 735 impurities, were excluded. We removed non-primary studies, such as editorials, chapters, thesis, reviews, and reports, approximately 1195 studies.

Then, we applied the inclusion and exclusion criteria based on the PICOC strategy: semantic interoperability, medical record (variants such as health record, medical record, patient record, hospital record), and standard.

Fig. 1 This graph presents, distributed by year of publication, the corpus of articles published during the range 2010 to September 2020



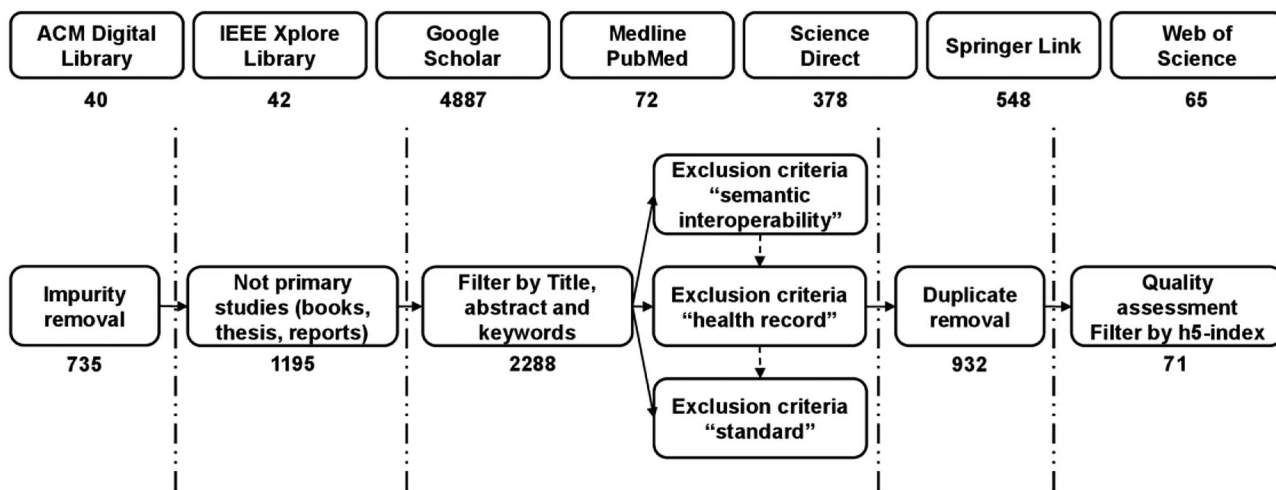


Fig. 2 The figure presents the entire selection process of the studies across the inclusion/exclusion criteria and quality assessment to conduct this systematic review

Finally, the inclusion and exclusion criteria are applied to the remaining studies to filter articles directly related to the research topic, totaling 2,288 excluded studies. The filters allow selecting studies according to defined terms.

5.3 Quality assessment

The last step had two parts; first, we filtered the studies by main interest area and evaluated the remaining corpus about its objectives. Many studies satisfied the inclusion and exclusion criteria but differed from the review’s interest topics. Thus, the quality assessment, the last step of the protocol, provides a cutoff parameter, where we look for studies published in relevant journals for the area of interest. Among these, we selected the highlighted studies. For this, we use

the h5-index metric, which quantifies the relevance of newspapers and conferences in the last five years, a Google metric and works with the highest number H [24]. Some studies that presented relevant discussions but did not appear in this index were separated to contribute to our discussion section.

In the final selection, some bases showed a predominance of accepted articles. For example, ScienceDirect with eight studies followed for PubMed Medline with seven and Google Scholar with six studies. Next was SpringerLink with four and IEEE Explore with two, while ACM Digital Library and Web of Science had one from each research base. Figure 3 present the final corpus articles accepted in this systematic review.

Figure 3 shows the accepted articles distributed by year. Table 3 shows where each article was published (journal

Fig. 3 This graph presents the articles accepted after the selection process, showing the number of articles by published year

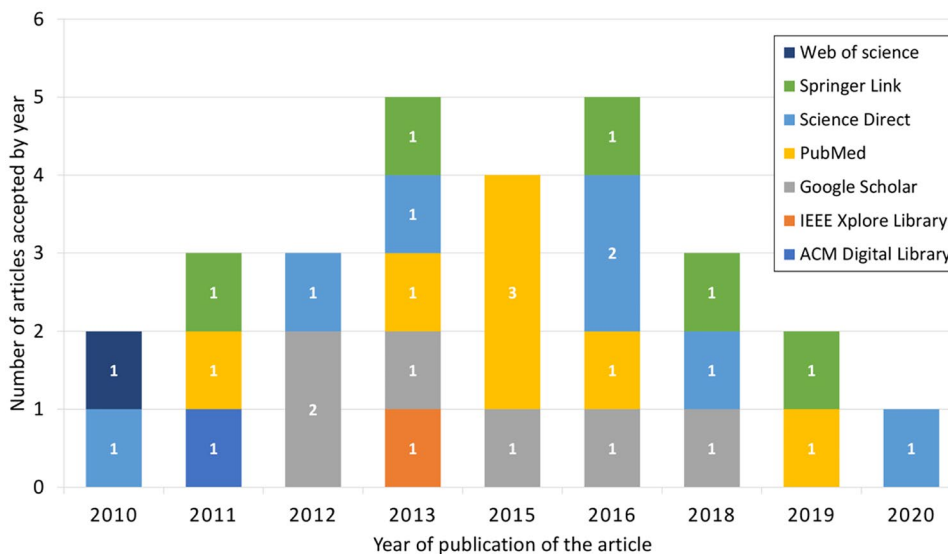


Table 3 This graph presents the number of papers by year of publication after finishing the selection and filtering steps

Journal / Conference	H5-index	Amount
ACM International Conference on Information and Knowledge Management	54	1
BioMed Research International	97	1
BMC Medical Informatics and Decision Making	41	4
BMC Research Notes	44	1
Computer Methods and Programs in Biomedicine	55	2
Future Generation Computer Systems	86	1
Hawaii International Conference on System Sciences	45	1
IEEE Journal of Biomedical and Health Informatics	69	1
International Conference on Autonomous Agents and Multi-Agent Systems	40	1
International Journal of Medical Informatics	55	2
Journal of Biomedical Informatics	60	5
Journal of Medical Internet Research	96	1
Journal of Medical Systems	58	1
Journal of the American Medical Informatics Association	67	4
Knowledge-Based Systems	85	1
Procedia Technology	34	1

or congress), the number of articles per location, and the H5 index used in the quality assessment stage. Three times the journals had an above-average acceptance, a scenario explained by the applied area, which is related to the research question of this review. We highlight the International Journal of Medical Informatics, with five accepted studies, BMC Medical Informatics and Decision Making, and the Journal of the American Medical Informatics Association, with four accepted studies for each, a relevant number compared to the rest of the studies shown in Table 3 below.

5.3.1 Data extraction

The information extracted from the articles aims to answer the question of interest in the review and identify approaches and solutions developed that allow achieving semantic interoperability in health records. Table 4 presents the final list of articles selected for this systematic review.

Follows the analysis of the selected studies against the questions of interest, and each other answered individually.

5.3.2 SQ1 – What are the health standards adopted in the studies?

There is no consensus on a global standard for electronic health records, and the studies selected for this review reinforced this scenario. However, the extracted data shows a trend in the standard choice with a multilevel approach, as such openEHR, ISO/CEN 13,606, and HL7 formats. That dual model approach allows specialists in health and technology to perform in a joint work. Most of the studies had related advances toward a semantic dataset choosing

standards to achieve semantic interoperability, as shown in Table 5.

The information extracted from studies demonstrates the open health standards as a trend, especially towards the two-level—openEHR and ISO 13606. Besides exchanging data and ensuring semantic interoperability concerns, as in [46], the authors had developed a framework combining ontology resources to predict high-risk situations in pregnancy. Additionally, the article contributed a summary analysis of 3 open standards, openEHR, ISO 13606, and HL7 CDA, showing their advantages and disadvantages. On the other hand, the studies [3, 4, 42] seek solutions for using openEHR and ISO13606 jointly, two open standards and similar definitions. This approach to approximation of the standards would allow the normalization of data. Although both standards use the ADL (Architecture Description Language), several differences in their types and definitions still need harmonizing.

In [16, 31, 43], the authors had explored slightly different opportunities from the known target of standards. For instance, [16] presented a methodology to represent the dependencies among data elements, concepts, and archetypes on a three-level Bayesian network and used the inference process to discover relevant archetypes—with promising results against the traditional search platform CKM. On the other hand, the authors in [43] improved patients' current post-sale drug surveillance process. The data come from voluntary reports (spontaneous reports, yet just about adverse incidents). The EHR adoption would allow tracing a complete patient medical history and to predict potential risk factors.

The authors in [31] also explored a different opportunity. They developed a federated Metadata Registry/Repository (MDR) – a metadata database of data combining Common

Table 4 This table presents the first author and respective studies by year, relating the publisher and the kind of place it was published

First Author	Ref	Year	Publisher	Type
Catalina Martínez-Costa et al.	[4]	2010	Elsevier	Journal
Georg Duftschmid et al.	[25]	2010	Elsevier	Journal
Adel Taweel et al.	[15]	2011	ACM	Journal
Bernard de Bono et al.	[26]	2011	BioMed Central	Journal
Leonardo Lezcano et al.	[27]	2011	Elsevier	Journal
David Mendes et al.	[28]	2012	Elsevier	Conference
Marcos Menárguez-Tortosa et al.	[29]	2012	Springer	Journal
Visara Urovi et al.	[30]	2012	IFAMAS	Conference
Anil Sinaci et al.	[31]	2013	Elsevier	Journal
Arshdeep Bahga et al.	[32]	2013	IEEE	Journal
Carlos Sáez et al.	[33]	2013	Elsevier	Journal
Georg Duftschmid et al.	[34]	2013	BioMed Central	Journal
Jesualdo Tomás Fernández-Breis et al.	[35]	2013	Oxford Academic	Journal
Carlos Marcos et al.	[36]	2015	Oxford Academic	Journal
Catalina Martínez-Costa et al.	[37]	2015	Oxford Academic	Journal
Luis Marco-Ruiz et al.	[38]	2015	Elsevier	Journal
María del Carmen Legaz-García et al.	[39]	2015	Oxford Academic	Journal
Afef Samet Ellouze et al.	[40]	2016	Elsevier	Journal
Hans Demski et al.	[41]	2016	BioMed Central	Journal
María del Carmen Legaz-García et al.	[42]	2016	Elsevier	Journal
Mustafa Yuksel et al.	[43]	2016	Hindawi	Journal
Raimundo Lozano-Rubí et al.	[44]	2016	Elsevier	Journal
Lingtong Min et al.	[45]	2018	BioMed Central	Journal
Mário et al.	[46]	2018	Elsevier	Journal
Shellon Blackman	[47]	2018	IEEE	Conference
Spyridon Kalogiannis et al.	[48]	2019	BioMed Central	Journal
Lin Yang et al.	[16]	2019	JMIR	Journal
José Alberto Maldonado et al.	[3]	2020	Elsevier	Journal

Data Elements (CDE) and HL7 CCD (Continuity of Care Document) models, proposing extensions to ISO 11179. Moreover, it was implemented in [36] the Health Level 7 (HL7) Virtual Medical Record (vMR) as a component service-based, that aims to collect patient data from different databases to allow the use in EHR data clinical decision support as a gateway between data sources and components.

5.3.3 SQ2 – What are the terminologies or health repositories used?

Terminologies and vocabularies can be understood as extensive collections of terms for a knowledge domain, making the language common. Terminologies aim to prevent local expressions, neologisms, and human typing from entering the EHR, thereby adopting formal classifications, e.g., diseases, events, procedures, specimens. Healthcare institutions share sensitive information, and systems must convey and not miss the meaning. One way possible is to build a local repository and manage an environment that applies proprietary concepts. However, the adoption of international terminologies ensures

that using global terms and other classifications will have the same meaning on the other side – to any receiver. Table 6 present the terminologies usually adopted by the studies.

Health standards generally easily adhere to different terminology as they are sets of data that can be entered into the repository and accessed and updated. Table 6 shows some terminologies most used by the selected studies. We highlight SNOMED-CT, the most cited terminology, bringing a clinically validated, semantically rich, controlled vocabulary. That facilitates evolutionary growth in expressivity to meet emerging requirements [50]. Also was often quoted and discussed, ICD is a diagnostic classification standard for clinical and research purposes. It defines a universe of diseases, disorders, injuries, and related health conditions [51], and in 2022 will publish the new version, the ICD-11.

While the adoption of standards is essential to achieve interoperability in EHR and effectively exchange information between different providers, the next level requires sharing knowledge and adding semantic value. A common language makes information comprehensible among who sent and received, allowing inferences in data and creating new

Table 5 Health standards used in the selected studies

Health standard	Reference
CEM (Clinical Element Models)	[39]
DICOM (Digital Imaging and Communications in Medicine)	[26]
HL7 CCR (Continuity of Care Record)	[15]
HL7 CDA (Clinical Document Architecture)	[33, 43]
HL7 CDD (Continuity of Care Document)	[31, 43]
HL7 HQMF (Health Quality Measure Format)	[43]
HL7 V2 and V3	[32]
HL7 VMR (Virtual Medical Record)	[36]
IHE (Integrating the Healthcare Enterprise)	[30]
ISO 13606	[3, 4, 25, 29, 42, 44]
Master data standardization and translation (MDST) model	[47]
openEHR	[3, 4, 16, 27, 32, 34, 37–42, 45, 46]

<http://www.clinicalelement.com/>
<https://www.dicomstandard.org/current>
<http://www.hl7.org/implement/standards/>
<https://www.ihe.net/>
<http://www.en13606.org/information.html>
<https://www.openehr.org/>
<https://ckm.openehr.org/ckm/>

connections from existing data. According to [49], semantic modeling essentially means linking words and terms to their senses, which is your main challenge.

Adopting a terminology engages more than one HIMSS level, as it condenses structural decisions regarding the system design, team, and organizational choices. The institution's medical staff must accept adherence, use the terminologies, and publicly reflect that choice. After compliance around the organization and team, the information conforms to a standard and shares it with other providers.

5.3.4 SQ3 – What are the approaches used?

The selected studies reinforce the advantages of adherence to standards for achieving semantic interoperability. Exploring new approaches involving technologies and health standards to extract better results from consolidated standards has shown growth. Likewise, some studies using the use of semantic web technologies to meet information extraction and harmonization demands. Table 7 present the principal purpose of the selected studies.

Table 6 The following are the international terminologies and classifications applied by the studies aiming at a shared vocabulary focusing on keeping the real meaning

Name	Reference
ICD-9 (International Classification of Diseases-version 9)	[43, 46]
ICD-10 (International Classification of Diseases-version 10)	[43]
SNOMED-CT (SNOMED Clinical Terms)	[27, 28, 31, 33, 36, 37, 40, 42–44]
LOINC (Logical Observation Identifiers Names and Codes)	[3, 36]
MeSH (Medical Subject Headings)	[42]
MedDRA (Medical Dictionary for Regulatory Activities)	[43]
WHO-ATC (World Health Organization Anatomical Therapeutic Chemical)	[43]
ChEBI (Chemical Entities of Biological Interest)	[26]
FMA (Foundational Model of Anatomy Ontology)	[26, 28]
BTL2 (BioTopLite2)	[37]
BioTop	[28]
Relation Ontology	[28]
CDE (Common Data Element)	[31]

Table 7 The studies had more than one objective, so this table separates them into applications and proposed approaches to meet interoperability demands in health records, categorizing them according to the principal approach to developing the contribution

Approach to develop the contribution	Reference
An ontology and rules-based clinical models mapping approach to solving classification/recommendation problems in Clinical Decision Support System	Rules: [27, 33] Ontological: [27, 46]
An ontological-based representation (a common mediator among clinical data, terminologies, and clinical models) or semantic web-based	Ontology as mediator: [4, 15, 30, 37, 39, 40, 43, 47] Other semantic web tools [28, 31, 42–44]:
An ontological-based representation (clinical models and clinical data)	Clinical models: [15, 27, 42] Clinical data: [28, 42]
Workflow approach as Cloud-based (component-based architecture) or Service-based framework	[3, 32, 36, 41]
Bayesian network-based approach to retrieve clinical artifacts (archetypes)	[16]
A collaborative framework to create clinical models by domain experts (less technical information more domain knowledge)	[45]
Metadata standards-based framework to enrich artifacts with controlled terminologies and semantic labeling	[15, 26, 31, 32]
XML-based EHR extract framework to represent archetype structure and constraints in XML and XML schema	[25, 38]
Building automatically interfaces from archetypes by XML representations	[29, 34]
IHE-based approach jointly agent-based framework – a solution to exchange patient data between the community	[30]

Standards to allow semantic interoperability have been a common choice in electronic health records systems. The papers selected for this review strengthen that scenario and justify solving problems by developing solutions for this purpose. An initiative regarding making health records available for secondary use in [42, 43], also reinforces a consequent advantage of implementing semantic interoperability since the improved data quality is part of the process. These works are also strongly related to the application of representations of clinical models [15, 27, 42], OWL semantic structures, where they also serve as semantic mediators [15, 26, 31], whereas [30] also added an agent-based system to coordinate the community IHE.

Hundreds of biomedical ontologies are available in OWL format in the BioPortal repository, including many medical terminologies, which justified using ontologies in some studies to convert clinical models, data, and other terminologies to this representation format natural. According to the authors in [42], exploring data semantic representation in ontologies is justified because other structures usually have explicit connections between data, and ontological systems allow reasoning to build different meanings. For this reason, to represent EHR metadata into ontologies typically appear like a good choice for semantic goals.

Moreover, the use of ontologies has also proved promising for mapping scenarios of rules data access. As demonstrated in [30], an agent coordination infrastructure uses an OWL (Web Ontology Language) ontology to map the access rules of organizations from the community to the data. On the other hand, the authors in [28] developed an automatic extraction from semi-structured data using ontologies. They

represented those concepts into ontologies of clinical vocabularies—another successful use of ontologies.

In line with the sharing and reuse of existing clinical models, some studies propose creating automated interfaces based on archetypes from the *openEHR* and ISO13606 standards, such as [29, 34]. On the other hand, some studies have identified novel architectures of service involving workflows in cloud environments, aiming to enable a tool to set resource pipelines, as presented in the works [3, 32, 36, 41], to access health data. Table 8 introduces the semantic web technologies used by the studies.

Despite being commonly used for semantic representation, semantic web technologies have broad applicability. Some studies have explored the representation of archetypes, rules, and relationships between different reference models (RM) of the *openEHR* and ISO13606 standards. The authors in [40, 44] have represented the reference model and archetype constraints into OWL ontology, which describes the instances and allows maintaining only one information (removing duplicated cases) to keep a relationship between RM and archetype.

Exploring other advantages, [3, 42] had presented EHR data into RDF and OWL structures – these transformations through the use of the Semantic Web Integration Tool (SWIT), as shown previously in Table 8. Additionally, they chose a graph database; respectively, the first used Neo4J Graph Database, and the last chose Virtuoso. However, both studies allow using Linked Open Data (LOD).

The studies also showed a trend by combining health standards and semantic web technologies to achieve semantic

Table 8 The different semantic web technologies used in studies to solve semantic problems

Semantic web technologies	Reference
LOD (Linked Open Data)	[3, 31, 42]
OWL (Ontology Web Language)	[3, 4, 26, 31, 39, 40, 42, 44, 46]
OWL-DL (Ontology Web Language-Description Logic)	[27, 28, 30, 37]
OWL-S—(Ontology Web Language-Services)	[15]
RDF (Resource Description Framework)	[3, 26, 31, 43, 47]
SKOS (Simple Knowledge Organization System)	[31, 43]
SPARQL (Protocol and RDF Query Language)	[26, 30, 37, 43]
XQuery script	[3]
N3 (Notation 3)	[43]
SWRL (Semantic Web Rule Language)	[27]

interoperability. There was no consensus about ontologies type adoption—OWL and RDF ontology. Both are used to represent the reference model and archetypes into ontologies. As reinforced by [4], the Archetype Definition Language (ADL) usually represents archetypes and has more of a syntactic orientation. Thus, it has disadvantages for achieving semantic interoperability, justifying the combining ontologies and clinical models. Also, the work presented in [27] used ontologies combining Semantic Web Rule Language (SWRL), and these rules aimed to allow applying logical deductive reasoning between archetypes and terminologies used.

Table 9 Databases used on studies

Database	Description	Ref
Virtuoso	Data virtualization platform of knowledge graphs and there is an open-source version. Allows using SQL and SPARQL	[31, 42]
Neo4J	Native graph database and open source. It is prepared to use Cypher (a graph query language)	[3]
HBase	It is a Hadoop database, a distributed, scalable, big data store	[32]
ARM	A proprietary approach capable of generating relational databases using archetypes and templates for archetype based EHR systems	[45]
Jena TBD	A component to RDF store and allow query with SPARQL	[31]
BaseX	High-performance XML database engine and a highly compliant XQuery and open source	[36]
Oracle	A multi-model database	[34]
Think!EHR	A big-data, high-performance platform designed to store, manage, query, retrieve and exchange structured electronic health record	[38]
PostgreSQL	An open-source object-relational database	[30]

<https://virtuoso.openlinksw.com/>

<https://neo4j.com/>

<https://hbase.apache.org/>

<https://jena.apache.org/documentation/tdb/>

<https://basex.org/>

<https://www.oracle.com/index.html>

<http://www.marand.com/thinkehr>

<https://www.postgresql.org/>

This systematic review focuses on understanding the different approaches proposed to achieve semantic interoperability through health standards, and some studies have indicated promising results when selecting a specific database. Although there are unusual and justified databases in the studies as to the reason for selection, we also kept the most typical databases. Then, Table 9 present the database used by the studies.

The selected studies the not show a clear trend as to whether the health standard influences the choice of database, although it is possible to assess some characteristics. For example, the Virtuoso multi-model bank appeared in solutions using openEHR and HL7, not showing a dependency relationship with the standard type. However, a graph database allows ontologies querying if existent, reinforcing a possible hybrid solution with semantic web tools.

On the other hand, ARM and Think!EHR presented a framework designed to support archetype storage and allow querying in ADL. Therefore, there is a dependency of choice when using ISO 13606 and openEHR standards. The other storage structures did not show a strong relationship or dependence on health standards, and few studies discussed the type of database adopted by their solutions.

5.3.5 SQ4 – What are the main security concerns used?

One of the concerns sharing data is security, such as using anonymized data and methods for de-identification. Furthermore, the exchange between different organizations must

Table 10 The Table presents the usually followed guidelines inside academic and controlled environments to use data from the real world

Privacy concerns	Description	Reference
Anonymization on safe zone	To use the data, the data provider first anonymized all data in a safe zone inside the health institution	[28, 35, 38, 43]
User profiles and restrict access levels	According to institution agreement, researchers use the data through restricted user profiles with different access levels	[15, 42, 47]

encompass safe policies, even to exchange among proprietary systems — safety issues such as maintaining security and integrity without breaking the quality of patient data.

The studies used several kinds of data, with different sources, such as from patients coming from the real world, such as [46] pregnancy data, [3, 42] and [3] colorectal cancer screening protocols and lab tests, [43] pharmacoepidemiology data, [40] infants affected by Cerebral Palsy [45], coronary computed tomography angiography, [36] atrial fibrillation (AF), [33] diabetes mellitus, [37] chronic heart failure, [27] abnormal reactions and allergies patient's data. Also, other studies used only results from lab tests as [26] radiology data, [35] histopathology reports, [38] test results of pertussis and salmonella. On the other hand, there was data synthetically generated in [30–32] that prevents any concerns about privacy or security.

The selected studies did not show a constant concern about data protection laws. This characteristic may be a consequence of the scientific character of the research since the application scenarios are generally academic environments and controlled. We highlight the two definitions observed regarding privacy and security concerns. In [32], HIPAA (Health Insurance Portability and Accountability Act) /HITECH and HITECH, and some directives from ISO 13606 applied on [3, 46], to store and manage health data.

On the other hand, we can consider that by keeping the security policies independent of the interoperability solutions, the authors aim to make the solutions generic in terms of local laws. Each country has its rules regarding using, sharing, and storing personal data – identification, financial or medical history. Thus, by keeping research an independent decision, studies are concerned with carrying out experiments that reflect real needs using real-world data.

Health Insurance Portability and Accountability Act (HIPAA) aims to ensure rights to citizens in the U.S., such as access to health records and request a copy of their data. As a patient, he can ask to correct something wrong with its health data, safe strategies to share data, and much more. To achieve that, they applied some methods to ensure security as authentication (SSO), authorization (OAuth), identity management, securing data at rest (using 256-bit advanced encryption standard), data in transit (HTTP with SSL), and auditing using a log data on the access. The last security layer represents the EHR stored alone because that provides data without the patient's identity. Below,

we highlight some characteristics regarding privacy, which some studies have indicated as necessary measures for using real-world data in their studies.

Studies that indicated the use of real-world data demonstrated some data providers' concerns to enable the use of the information without compromising the patients. Therefore, we can observe that, regardless of the study, those who used real-world data need to apply some form of data anonymization. Table 10 present the two ways system providers usually allow data from real-world to academic researchers.

When the responsibility for data anonymization rests with the provider, the institution usually executes in a safe zone. Data only leave the health institution after being completely de-identified. On the other hand, the provider may allow researchers to access the system internally. In that case, the electronic registration system must have access levels for users, not allowing unrestricted access.

5.3.6 SQ5 – What are the evaluation approaches used?

The evaluation of an EHR involves storing data with quality, the end-users experience (health care professional), keeping complex information from specialist physicians, and ensuring the exchange of information between healthcare providers. Therefore, a scenario where the system must adapt to the routine of health professionals and facilitate data collection but meet the demands of sharing data with quality.

In this way, approaches to evaluation aim to identify measures to validate the different levels of interoperability (foundation, structural and semantic levels) that an EHR achieved and map the challenges of systematically doing it. The studies presented other evaluation methods, and Table 11 shows the different applied approaches.

Applying questionnaires as an evaluation method allows identifying which changes – according to the user and domain expert – would be more effective in improving the final solution through constant feedback. The author in [43] had an evaluation based on ISO/IEC 25,040 (SQuaRE), System Usability Scale (SUS), and Health IT Usability Evaluation Scale (Health-ITUES). On the other hand, functional tests allow the initial assessment to establish metrics to outcomes and automated tests.

However, none of the approaches assess the extent of interoperability or validate which different interoperability levels

Table 11 Follows we list the evaluation approaches and the types of data used in the articles

Evaluation method or metric	Article
Questionnaire (end-user feedback)	[16, 43]
Questionnaire (specialist feedback)	[16, 42, 43, 47]
Functional evaluation	[4, 15, 29, 34, 39, 40, 44, 45]
Case study (applied inside or outside health institution environment)	Outside [3, 27, 28, 30, 31, 33, 35, 37, 38, 42] Inside [26, 33, 36]
Comparative analyses (traditional data vs novel data)	[35, 43]
Unitary tests	[33]
Usability and compatibility across browsers	[29]

the applications have achieved – as per the levels defined by HIMSS [9]. Therefore, evaluating interoperability, as a result, is a challenge and an opportunity for further research, looking at more effective and unique methods.

5.3.7 SG1 – What is the state of art in health standards applied in health records?

There are some challenges to be faced by the health specialists and IT professionals to achieve semantic interoperability in EHR. This review identified different approaches to work around known problems and showed what technologies and strategies are being got in the area. Using the two-level (also called dual model) approach is not a global view of health standards. However, it can allow a common syntax and clinical data representation between the systems [38]. Furthermore, regardless of the management system adopted, this allows of making the clinical model software-independent: the specialized health professionals model the clinical document definition and terminologies as needed. In contrast, computer professionals are concerned with the structure, architecture, and choice of technologies to enable the use of knowledge defined by the health professional.

While there is no health standards consensus, we can see a trend in several studies using the two-level standard. This preference occurs because the OpenEHR standard provides a more mature public library – tools, archetypes, community support, and guides [34]. Furthermore, ISO 13606 is not publicly available, and developing novel archetypes to satisfy a system would take a lot of time [25]. Besides, to improve the adoption of the standards, many studies combined semantic web technologies and health standards with openEHR. A common practice identified in the articles was the structure representation of archetypes and reference models in ontologies, combined with rules mapping their constraints—another use involved ontologies representing metamodels with general information and ontologies for disease classification.

The solutions proposed by the selected articles brought different approaches to achieve semantic interoperability. In

this scenario, Fig. 4 presents a taxonomy proposal around semantic interoperability in health records. Although the taxonomy is organized into five categories, the articles usually combine multiple themes to achieve the semantic interoperability challenge, so they are not exclusive sections. However, in the process of analyzing the studies, we defined five main categories: 1) Health Standards, 2) Classification and Terminologies, 3) Semantic Web, 4) Data Storage, and 5) Evaluation.

A taxonomy may allow us to assess a broader scenario—e.g., what artifacts involve building an EHR semantically interoperable. We reinforce that the development of an EHR must have semantic interoperability as a goal from early project design. The decision to adopt standards and terminologies at the beginning can facilitate the development process, once the artifacts are discussed as an integral part of the project, as mandatory.

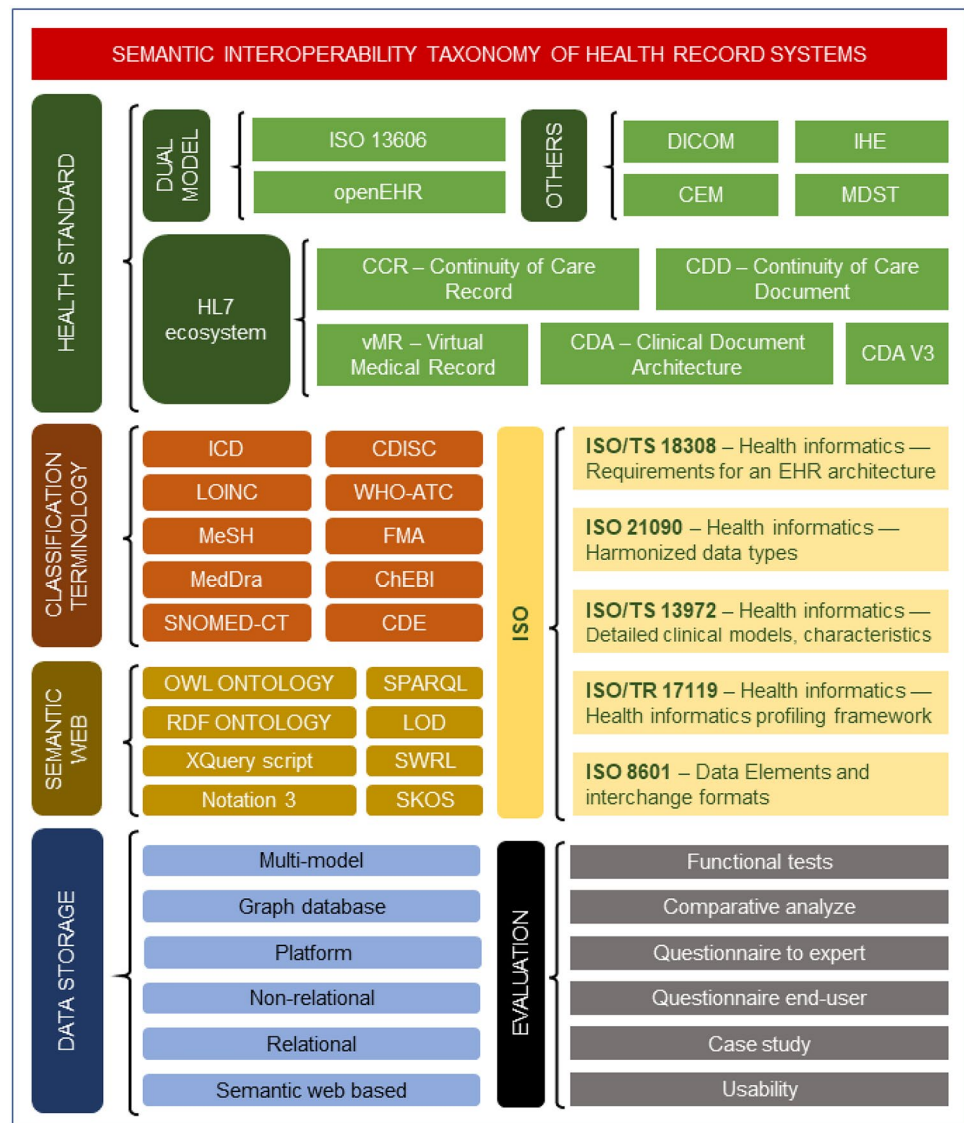
The taxonomy scope involved the accepted articles, all technologies, standards applied to solve semantic interoperability and exchange data across organizations and health systems. The selected studies described many tools usually used to solve interoperability problems. However, we focus on solving the article research question and add that to the taxonomy list. We highlighted the taxonomy is not an exhaustive map to all semantic interoperability-related tools, only applied to the selected studies.

The first category, Health Standard, shows the standards used by the studies, with three subcategories. The dual model—openEHR and ISO 13606 inside share this architecture. Likewise, the HL7 standards have an ecosystem of standards, however, not all have the same goal. The HL7 organization also maintains a comprehensive list of terminologies³ compatible with available standards.

The choice of health standard may be related to some implementation team characteristics. The HL7 standard prioritizes a friendly relationship with the developer, with

³ http://www.hl7.org/documentcenter/public/standards/vocabulary/vocabulary_tables/infrastructure/vocabulary/vocabulary.html

Fig. 4 Proposed taxonomy for semantic interoperability in health records



technical documentation and structures similar to the development ecosystem. In contrast, the openEHR proposes the opposite. Healthcare professionals have a user-friendly interface to create clinical models (in archetypes), focusing on defining knowledge. On the other hand, the artifacts defined for the openEHR standard—archetypes—have a structure that allows for in-depth detailing of clinical concepts, enabling interoperability at a semantic level.

Using standards inherently raises the concern of employing international terminologies to represent clinical terms and concepts. We show the vocabularies used in the Classification and Terminologies section of the taxonomy. Some terminologies have a massive adherence of studies, such as SNOMED and Logical Observation Identifiers, Names, and Codes (LOINC). Most health standards allow using more than one terminology in the same clinical document with semantic binding once a healthcare organization has legacy

data and treats data through many vocabularies—enabled in openEHR and HL7, consequently in IHE.

Adopting terminologies as an inherent part of the EHR brings benefits, such as the standardization of everyday expressions, ensuring semantic contextualization concerning diseases, adverse events, and general classifications. The semantic contextualization allows new connections to collected data once the structure—patient’s history, lab tests, exams—has a semantic binding through international vocabularies. Unfortunately, there are still challenges when discussing the patients’ history. Usually have an open and unstructured text field, a format more accessible to physicians and health professionals. However, an unstructured text field does not readable to the machines.

The studies treat the unstructured data using semantic tools to extract and structure these data. Although there is a low variety of semantic web technology, the combination

with health standards was almost unanimous. The most used tools often follow the definitions established by the W3C, such as OWL, RDF, SPARQL, SKOS, as reinforced in the recent research field review [52]. The Semantic Web section of the taxonomy shows the ontologies most used, such as SKOS, OWL, OWL-DL, and RDF as final semantic resources.

Although the studies do not explore the benefits of the choice of storage solution, some storage solutions are related to the type of health standard adopted. The trend towards the use of ontologies can impact the choice of storage solution since it is interesting to allow querying the ontological structure using SPARQL and exploring reasoning. In this scenario, Virtuoso, Neo4J, and Oracle databases present solutions that adequately meet ontologies' storage demand.

In the Data Storage section, we map the different solutions used for storage to the taxonomy and categorize them according to the structure they provide. In addition, we highlight among these solutions two that specifically implement clinical document storage—archetypes such as ARM and Think!EHR. The other solutions use web-based graphical and semantic databases (W3C compliant), similar in their storage structure.

Finally, not all papers presented an evaluation format for their experiments or results obtained. Some authors had highlighted the use of case studies in a controlled environment to apply their final experiments. Often the authors used questionnaires with end-users to evaluate the user experience. The functional tests also had a prevalence, tracking specific modules with problematic scenarios while being developed to resolve before use. In some solutions, the authors described their partial evaluation methods during the development process. They wanted to evaluate tools' accuracy or performance and not precisely the final solution, allowing for semantic interoperability. However, those methods did not consider the taxonomy because they did not influence the final discussion of the evaluated solution.

5.3.8 SG2 – What are the challenge and open questions to semantic interoperability in health records?

The adoption of standards has grown one of the ways to ensure semantic interoperability in health record systems. However, there are many barriers to overcome in the health organizations, such as legacy data, semi-structured data, non-structured textual data, complex systems that are sometimes not compatible for exchanging information. Once get over those challenges yet have internal adversities as the concepts, and medical terms used across the organization must preserve that meaning externally shared.

The search for interoperability at a semantic level involves combining different vocabularies from different areas to maintain a common meaning. For this scenario, the use of

terminologies, ontologies, and global classifications plays an important role. However, since there is no worldwide consensus on health standards or unique clinical vocabularies, deciding which standard remains a technical issue when it should be an organizational one. Some research projects towards proposing vocabulary harmonization within EHR systems represent a promising alternative. These can allow the adoption of different vocabularies but with unique meanings. In other words, effective EHR communication depends on standardizing syntax, structure, and semantics (from the chosen architecture to the vocabulary used). The papers presented showed concerns about the difficulty in standardizing and normalizing data in legacy systems, as previously mentioned in [42, 43]. Complex models and legacy data are some of the limiters in secondary use and data reuse [42] and different purposes such as clinical research and decision support systems [38]. The lack of common terminologies and the extensive use of proprietary concepts inside EHR become the interoperability complex and requires a normalization process [43].

6 Discussion

This review aims to understand the current scenario of solutions employed to solve semantic interoperability in health records. The selected studies allowed evaluating the current strategies used to meet interoperability at a semantic level. In addition, we highlight how interoperability must happen at the organizational level since the difficulties of integrating, sharing, and exchanging health records across all the organization's sectors. Interoperability at the semantic level requires organizational involvement, ranging from institution management to the involved teams such as technology staff and healthcare professionals. Therefore, it is essential to consider the reality in health institutions, where valuing the patient in terms of quality in health is part of all processes and interactions, by delivering quality of care, economy, and efficiency.

The studies evaluated in this review proposed some alternatives involving adopting semantic web technologies combined with health standards to meet this demand. For example, the prevalence of ontologies used to represent reference models and clinical information models of the standard used, as shown in Table 7. Moreover, we observed a preference for adopting semantic web technologies recommended by World Wide Web Consortium (W3C), such as RDF, OWL, SPARQL, SKOS, as shown in Table 8.

On the other hand, studies that explored the two-level standards showed OWL ontology's predominant use, while studies that explored available HL7 standards showed RDF ontology's predominant use. Despite this, in both scenarios (HL7 and two-levels – openEHR and ISO13606), the studies advocate adopting ontologies as a suitable format to

explore the use of metadata and semantic relationships between clinical terms of the collected data. The justification for using ontologies is regarded with the possibilities of inferences and rules, as the main reason. Also is important the obtained semantic orientation since other formats have a syntactic orientation.

Also, the studies selected highlighted a scenario of fragility, exposing the difficulties for an interoperable electronic medical record when in the fragile reality of health institutions. The existing, legacy, and sometimes non-structured data do not have any standard or vocabulary. Moreover, since there is no integrated system, the patient history is fragmented, with duplicate entries. This scenario is also aggravated at the level of system suppliers, maintaining proprietary structures for storage, which makes it impossible to share data with other institutions without an extended period for adjustments.

The situation within healthcare organizations is still far from the controlled environment often found in research projects. For this reason, it is necessary to observe reality and propose possible and viable alternatives. In this fragile and restricted scenario of institutions, as approached by HIMSS [9], the systems it partially reaches the first two layers of interoperability—structural and fundamental. Moreover, it is necessary to use an international standard to guarantee a widely accepted syntax and formats even to achieve just these two layers. This is still a significant challenge.

That scenario is reinforced by [11, 20], regarding the three fundamental levels of interoperability to improve the workflows across health information systems to allow true interoperability. Inside healthcare institutions, the collect data happens in multiple formats, tabulated and structured—when well done—however essentially as textual non-structured, free form, collected from health professionals in fragmented contexts, added of the lack terminologies and without any international denomination to concepts.

We identified some open questions from this scenario that future studies can explore. For example, open health standards and international terminology knowledge are still far from extensive use in the market and are only widespread in academia (educational institutions) and large software providers. Therefore, the opportunity arises to contribute with measures that mitigate the recurring difficulties of standardization, making knowledge about available resources accessible, easy to implement, and providing adequate materials to facilitate the study.

Since knowledge about patterns is not widespread, few systems implement terminology, classifications, and dictionaries, as they are not part of the healthcare reality in their designs. Therefore, it is crucial to work on initiatives to promote the gradual adoption of terminologies in daily use. That lack means opportunities to propose methodologies to adopt a health standard, mapping possible difficulties and how to

face them, offering viable directives for the reality of health institutions.

The interoperability as a focus of integration and information exchange is part of the scope of software companies, a commonly raised concern. However, healthcare systems consider semantic interoperability a luxury when it should be a prerequisite since it directly implies the quality of the stored data, which will later use in secondary studies. Therefore, encouraging the construction of systems aiming to achieve semantic interoperability would allow exploring secondary use of EHR, data reuse, and clinical models and doing clinical research within institutions a reality.

7 Conclusion

This article presents an overview of international health standards usually applied to allow semantic interoperability in health records. We conducted a Systematic Literature Review based on the protocol proposed by Kitchenham [22]. Our research set involved seven scientific databases, from which 6032 studies were selected. After the application of the inclusion and exclusion criteria, quality assessment, the result was 28 accepted articles. These were used for complete reading and analysis according to the interest questions.

We observe that the predominance in adopting two levels (ISO13606 and openEHR) is primarily due to an open standard's nature, extensive documentation available, and is inherently designed to make use accessible to healthcare professionals. That accessibility allows them to build the necessary clinical models without the assistance of specialized professionals. It also refers to the level of granularity that the pattern allows representing since the archetypes can include deeply detailed levels of information regarding a specific clinical concept, allowing for a contextualized semantic level. Finally, it is possible to create a collection of archetypes to construct a template.

In general, we observed studies applied in medical records of hospitals, some scenarios using laboratory data and some experiments aimed at the exploration of clinical models (ISO13606 and openEHR standard). We also analyzed options to maximize the characteristics proposed in the models, such as recovery, forms semantic representation, techniques for enriching semantic relations through metadata and ontology standards. The conducted study made it possible to identify a growing concern about the adoption of open data models to represent clinical knowledge. The main standards mapped were openEHR, ISO13606, and the HL7 framework.

This review reinforces a promising scenario for exploration since there is not an international standard or global consensus on approaches to be adopted. Recent research has shown efforts to define a universal model involving the four levels of interoperability as a guide for a definitive directive.

Some studies aimed to allow secondary data exploration, such as clinical research or decision support systems. Ontologies have been widely used, and we observed good results in adopting semantic web technologies, mainly using ontologies combined with patterns, to increase data representation in formats with a semantic focus. With data represented in ontologies, this scenario encourages the exploration of linked databases (LOD) and databases based on graphics (Neo4J and Virtuoso), but without a final definition of the advantages of adopting these databases. Finally, we highlight the trend in the adoption of semantic web technologies recommended by the W3C. The advantages of health standards (clinical models and reference models) with terminologies combined with graph structures (ontologies) for representing data from electronic health records, relation, and constraints rules.

Acknowledgements and funding

The authors would like to thank the Coordination for the Improvement of Higher Education Personnel—CAPES (Financial Code 001) and the National Council for Scientific and Technological Development—CNPq (Grant number 309537/2020–7) for supporting this work.

Author contributions All authors have made a substantial, direct, intellectual contribution to this study.

Data availability Not applicable.

Code availability (software application or custom code) Not applicable.

Declarations

Conflict of Interest None.

References

- Kopanitsa G. Integration of hospital information and clinical decision support systems to enable the reuse of electronic health record data. *Methods Inf Med*. 2017;56(3):238–47. <https://doi.org/10.3414/ME16-01-0057>.
- Roehrs A, da Costa CA, Righi R da R, Rigo SJ, Wichman MH. Toward a Model for Personal Health Record Interoperability. *IEEE j biomedical health informatics* 2019;23(2):867–873. <https://doi.org/10.1109/JBHI.2018.2836138>
- Maldonado JA, Marcos M, Fernández-Breis JT, Giménez-Solano VM, Legaz-García M del C, Martínez-Salvador B. CLIN-IK-LINKS: A platform for the design and execution of clinical data transformation and reasoning workflows. *Comp Methods Programs Biomed* 2020;197:105616. <https://doi.org/10.1016/j.cmpb.2020.105616>
- Martínez-Costa C, Menárguez-Tortosa M, Fernández-Breis JT. An approach for the semantic interoperability of ISO EN 13606 and OpenEHR archetypes. *J Biomed Inform*. 2010;43(5):736–46. <https://doi.org/10.1016/j.jbi.2010.05.013>.
- Kim D-Y, Joshi KP. A Semantically Rich Knowledge Graph to Automate HIPAA Regulations for Cloud Health IT Services. 2021.
- Benson T, Grieve G. Why Interoperability Is Hard. In: *Principles of Health Interoperability*, 4th ed. Springer, Cham. 2021;21–40.
- Alexopoulos P. Mind the Semantic Gap. In: *Semantic Modeling for Data*, 1st ed. O'Reilly Media, Inc. 2020:3–12.
- Interoperability WP. *ACM Comput Surv*. 1996;28(1):285–7. <https://doi.org/10.1145/234313.234424>.
- HIMSS. Healthcare Information and Management Systems Society. What is Interoperability?. <https://www.himss.org/resources/interoperability-healthcare>. Accessed April 27, 2021.
- Gansel X, Mary M, van Belkum A. Semantic data interoperability, digital medicine, and e-health in infectious disease management: a review. *Eur J Clin Microbiol Infect Dis*. 2019;38(6):1023–34.
- Hayman JA, Dekker A, Feng M, et al.. Minimum Data Elements for Radiation Oncology: An American Society for Radiation Oncology Consensus Paper. *Pract Radiat Oncol*. 2019;9(6):395–401. <https://doi.org/10.1016/j.prro.2019.07.017>.
- Angula N, Dlodlo N. Towards a framework to enable semantic interoperability of data in heterogeneous health information systems in Namibian public hospitals. In: *Advances in Intelligent Systems and Computing*, vol 721. Springer Verlag. 2018;835–845.
- ISO C técnico. ISO 18308:2011(en), Health informatics — Requirements for an Electronic Health Record Architecture. 2011. <https://www.iso.org/obp/ui/#iso:std:iso:18308:ed-1:v1:en>. Accessed March 22, 2021.
- ISO/TR 20514. ISO/TR 20514: Health informatics — Electronic health record — Definition, scope, and context. 2005.
- Taweel A, Speedie S, Tyson G, Tawil ARH, Peterson K, Delaney B. Service and Model-Driven Dynamic Integration of Health Data. In: *Proceedings of the First International Workshop on Managing Interoperability and Complexity in Health Systems*. New York, NY, USA: Association for Computing Machinery. 2011;11–17.
- Yang L, Huang X, Li J. Discovering clinical information models online to promote interoperability of electronic health records: A feasibility study of openEHR. *J Med Internet Res*. 2019;21(5). <https://doi.org/10.2196/13504>.
- ISO 13606. ISO 13606:2019 Standard - EHR Interoperability. 2019.
- Salomi MJA, Claro PB. Adopting Healthcare Information Exchange among Organizations, Regions, and Hospital Systems toward Quality, Sustainability, and Effectiveness. *Technol Invest*. 2020. https://www.scirp.org/html/3-9900884_101997.htm Accessed.
- Gagalova KK, Elizalde MAL. What You Need to Know Before Implementing a Clinical Research Data Warehouse: Comparative Review of Integrated Data Repositories in Health Care *JMIR Formative*. 2020. https://formative.jmir.org/2020/8/e17687/?utm_source=TrendMD&utm_medium=cpc&utm_campaign=JMIR_Formative_Research_TrendMD_1 Accessed.
- Lee AR, Kim IK, Lee E. Developing a Transnational Health Record Framework with Level-Specific Interoperability Guidelines Based on a Related Literature Review. *Healthcare*. 2021;9(1). <https://doi.org/10.3390/healthcare9010067>.
- Adel E, El-Sappagh S, Barakat S, Elmogy M. Ontology-based electronic health record semantic interoperability: A survey. *U-Healthcare Monitoring Sys* 2019;315–352. <https://doi.org/10.1016/B978-0-12-815370-3.00013-X>.
- Kitchenham B, Brereton P. A systematic review of systematic review process research in software engineering. *Inf Softw Technol*. 2013;55(12):2049–75. <https://doi.org/10.1016/j.infsof.2013.07.010>.
- Roehrs A, Da Costa CA, Da Rosa Righi R, De Oliveira KSF. Personal health records: A systematic literature review. *J Med Internet Res* 2017;19(1). <https://doi.org/10.2196/jmir.5876>.
- Google scholar metrics. 2021. https://scholar.google.com/citations?view_op=top_venues&hl=en. Accessed December 16, 2020.
- Dufts Schmid G, Wrba T, Rinner C. Extraction of standardized archetyped data from Electronic Health Record systems based on the Entity-Attribute-Value Model. *Int J Med Informatics*. 2010;79(8):585–97. <https://doi.org/10.1016/j.ijmedinf.2010.04.007>.
- De Bono B, Hoehndorf R, Wimalaratne S, Gkoutos G, Grenon P. The RICORDO approach to semantic interoperability for

- biomedical data and models: Strategy, standards and solutions. *BMC Res Notes* 2011;4. <https://doi.org/10.1186/1756-0500-4-313>.
27. Lezcano L, Sicilia M-A, Rodríguez-Solano C. Integrating reasoning and clinical archetypes using OWL ontologies and SWRL rules. *J Biomed Inform.* 2011;44(2):343–53. <https://doi.org/10.1016/j.jbi.2010.11.005>.
 28. Mendes D, Rodrigues IP, Rodriguez-Solano C, Baeta C. Enrichment/Population of Customized CPR (Computer-based Patient Record) Ontology from Free-text Reports for CSI (Computer Semantic Interoperability). *Procedia Technol.* 2012;5:753–62. <https://doi.org/10.1016/j.protcy.2012.09.083>.
 29. Menárguez-Tortosa M, Martínez-Costa C, Fernández-Breis JT. A generative tool for building health applications driven by ISO 13606 archetypes. *J med sys* 2012;36(5). <https://doi.org/10.1007/s10916-011-9783-1>.
 30. Urov V, Olivieri AC, Bromuri S, Fornara N, Schumacher MI. An Agent Coordination Framework for IHE based Cross-Community Health Record Exchange. In: Moreno A, Isern D, editors. International conference on Autonomous Agents and Multiagent Systems (AAMAS). Valencia: Spain; 2012. p. 29–39.
 31. Sinaci AA, Laleci Erturkmen GB. A federated semantic metadata registry framework for enabling interoperability across clinical research and care domains. *J Biomed Inform.* 2013;46(5):784–94. <https://doi.org/10.1016/j.jbi.2013.05.009>.
 32. Bahga A, Madiseti VK. A cloud-based approach for interoperable electronic health records (EHRs). *IEEE J Biomed Health Inform.* 2013;17(5):894–906. <https://doi.org/10.1109/JBHI.2013.2257818>.
 33. Sáez C, Bresó A, Vicente J, Robles M, García-Gómez JM. An HL7-CDA wrapper for facilitating semantic interoperability to rule-based Clinical Decision Support Systems. *Comput Methods Programs Biomed.* 2013;109(3):239–49. <https://doi.org/10.1016/j.cmpb.2012.10.003>.
 34. Duftschmid G, Chaloupka J, Rinner C. Towards plug-and-play integration of archetypes into legacy electronic health record systems: The ArchiMed experience. *BMC Med Informatics and Decision Making.* 2013;13(1). <https://doi.org/10.1186/1472-6947-13-11>.
 35. Fernández-Breis JT, Maldonado JA, Marcos M, et al.. Leveraging electronic healthcare record standards and semantic web technologies for the identification of patient cohorts. *J Am Med Inform Assoc.* 2013;20(e2):e288–96. <https://doi.org/10.1136/amiajnl-2013-001923>.
 36. Marcos C, González-Ferrer A, Peleg M, Cavero C. Solving the interoperability challenge of a distributed complex patient guidance system: a data integrator based on HL7's Virtual Medical Record standard. *Journal of the American Medical Informatics Association : JAMIA.* 2015;22(3):587–99. <https://doi.org/10.1093/jamia/ocv003>.
 37. Martínez-Costa C, Cornet R, Karlsson D, Schulz S, Kalra D. Semantic enrichment of clinical models towards semantic interoperability. The heart failure summary use case. *J Am Med Informatics Assoc: JAMIA.* 2015;22(3):565–576. <https://doi.org/10.1093/jamia/ocu013>.
 38. Marco-Ruiz L, Moner D, Maldonado JA, Kolstrup N, Bellika JG. Archetype-based data warehouse environment to enable the reuse of electronic health record data. *Int J Med Informatics.* 2015;84(9):702–14. <https://doi.org/10.1016/j.ijmedinf.2015.05.016>.
 39. Legaz-García MD, Menárguez-Tortosa M, Fernández-Breis JT, Chute CG, Tao C. Transformation of standardized clinical models based on OWL technologies: from CEM to OpenEHR archetypes. *J Am Med Informatics Assoc: JAMIA.* 2015;22(3):536–544. <https://academic.oup.com/jamia/article-abstract/22/3/536/776073> Accessed.
 40. Ellouze AS, Bouaziz R, Ghorbel H. Integrating semantic dimension into openEHR archetypes for the management of cerebral palsy electronic medical records. *J Biomed Inform.* 2016;63:307–24. <https://doi.org/10.1016/j.jbi.2016.08.018>.
 41. Demski H, Garde S, Hildebrand C. Open data models for smart health interconnected applications—the example of openEHR. *BMC Med Inform Decis Mak.* 2016;16(1):137. <https://doi.org/10.1186/s12911-016-0376-2>.
 42. del Carmen L-G, Martínez-Costa C, Menárguez-Tortosa M, Fernández-Breis JT. A semantic web based framework for the interoperability and exploitation of clinical models and EHR data. *Knowl-Based Syst.* 2016;105:175–89. <https://doi.org/10.1016/j.knosys.2016.05.016>.
 43. Yuksel M, Gonul S, Erturkmen GBL, et al.. An interoperability platform enabling reuse of electronic health records for signal verification studies. *Stud BioMed res int.* 2016. <https://doi.org/10.1155/2016/6741418>.
 44. Lozano-Rubí R, Muñoz Carrero A, Serrano Balazote P, Pastor X. OntoCR: A CEN/ISO-13606 clinical repository based on ontologies. *J Biomed Inform.* 2016;60:224–33. <https://doi.org/10.1016/j.jbi.2016.02.007>.
 45. Min L, Tian Q, Lu X, An J, Duan H. An open EHR based approach to improve the semantic interoperability of clinical data registry. *BMC Med Informatics and Decision Making* 2018;18. <https://doi.org/10.1186/s12911-018-0596-8>.
 46. Moreira MWL, Rodrigues JJPC, Sangaiah AK, Al-Muhtadi J, Korotav V. Semantic interoperability and pattern classification for a service-oriented architecture in pregnancy care. *Futur Gener Comput Syst.* 2018;89:137–47. <https://doi.org/10.1016/j.future.2018.04.031>.
 47. Blackman-Lees S. Towards a Conceptual Framework for Persistent Use: A Technical Plan to Achieve Semantic Interoperability within Electronic Health Record Systems. In: Proceedings of the 51st Hawaii International Conference on System Sciences. scholarpace.manoa.hawaii.edu. 2018;10.
 48. Kalogiannis S, Deltouzos K, Zacharaki EI, et al.. Integrating an openEHR-based personalized virtual model for the ageing population within HBase. *BMC Med Inform Decis Mak.* 2019;19(1):25. <https://doi.org/10.1186/s12911-019-0745-8>.
 49. Alexopoulos P. *Semantic Modeling for Data*. books.google.com. 2020.
 50. SNOMED International. SNOMED International. <https://www.snomed.org/>. Accessed May 10, 2021.
 51. Organization WH. Classification of Diseases (ICD). <https://www.who.int/standards/classifications/classification-of-diseases>. Accessed May 10, 2021.
 52. Hitzler P. A Review of the Semantic Web Field. *Commun ACM.* 2021;64(2):76–83. <https://doi.org/10.1145/3397512>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.