J. Health Inform. 2023 Abril-Setembro; 15(2): 61-69

Artigo de Revisão

DOI: https://doi.org/10.59681/2175-4411.v15.i2.2023.993

Semantics in EHR for pediatric oncology: an integrative review

Semântica em Prontuários Eletrônicos para oncologia pediátrica: uma revisão integrativa

Semántica en la Historia Clínica Electrónica para oncología pediátrica: una revisión integradora

Elaine Barbosa de Figueiredo[1], Ferrucio de Franco Rosa[2], Ricardo Antônio Zanetti[3], Mariangela Dametto[4], Rodrigo Bonacin[5]

Keywords: Electronic Health Records, Biological Ontologies, Medical Oncology, Controlled Vocabulary

ABSTRACT

Objective: This study aims to analyze the use of Knowledge Organization Systems (KOS) as a means of enriching Electronic Health Records (EHR) for the domain of pediatric oncology. **Methods:** An integrative literature review method is applied. Three literature reviews were conducted, with a search for articles from 2016 to July/2023 in PubMed, Scopus, IEEE Xplore e ACM Digital Library written in English or Portuguese. **Results:** 52 articles were analyzed. The results point out the adopted standards for EHR specification and describe the most frequent KOS used with EHR in oncology as well as in the pediatric oncology domain. **Conclusion:** Although there are efforts to adopt international standards for EHR, several projects do not make use of these standards. EHR systems for oncology, in general, make wider use of KOS, while in pediatric oncology the focus is on those related to genetics. There is a need for further research to integrate KOS with international standards.

Descritores: Registros Eletrônicos de Saúde, Ontologias Biomédicas, Oncologia, Vocabulário Controlado

RESUMO

Objetivo: Este estudo tem como objetivo analisar o uso de Sistemas de Organização do Conhecimento (SOC) como meio de enriquecimento do Prontuário Eletrônico do Paciente (PEP) para o domínio da oncologia pediátrica. **Métodos:** Foi aplicado um método de revisão integrativa da literatura. Foram realizadas três revisões de literatura, com busca de artigos de 2016 até Julho/2023 em PubMed, Scopus, IEEE Xplore e ACM Digital Library escritos em inglês ou português. **Resultados:** Foram analisados 52 artigos. Os resultados apontam os padrões adotados para a especificação de PEP e descrevem os SOC mais frequentemente usados com PEP na oncologia e no domínio da oncologia pediátrica. **Conclusão:** Embora existam esforços para adotar padrões internacionais para PEP, vários projetos não fazem uso desses padrões. Os sistemas de PEPs para oncologia, em geral, fazem uso mais amplo de SOCs, enquanto na oncologia pediátrica o foco está nos relacionados à genética. Há necessidade de mais pesquisas para integrar PEP com padrões internacionais.

Descriptores: Registros Electrónicos de Salud, Ontologías Biológicas, Oncología Médica, Vocabulario Controlado

RESUMEN

Objetivo: Este trabajo tiene como objetivo analizar el uso de Sistemas de Organización del Conocimiento (SOC) como medio de enriquecimiento de la História Clínica Electrónica (HCE) para el dominio de la oncología pediátrica. **Método:** Se utilizó un método integrador de revisión de la literatura. Se realizaron tres revisiones bibliográficas, buscando artículos desde 2016 hasta julio/2023 en PubMed, Scopus, IEEE Xplore y ACM Digital Library. **Resultados:** Se analizaron 52 artículos. Los resultados apuntan a los estándares adoptados para la especificación de HCE y describen el SOC más frecuentemente utilizado con HCE en oncología y también en el campo de la oncología pediátrica. **Conclusión:** Si bien existen esfuerzos para adoptar estándares internacionales para HCE, varios proyectos no hacen uso de estos estándares. Los sistemas HCE para oncología, en general, hacen un mayor uso de los SOC, mientras que en oncología pediátrica el foco está en los relacionados con la genética. Se necesita más investigación para integrar HCE con estándares internacionales.

[1] *Professora, Serviço Nacional de Aprendizagem Comercial, São Paulo, SP, Brasil*

[2] *Pesquisador, Centro de Tecnologia da Informação Renato Archer, Campinas, SP, Brasil e Professor, Centro Universitário Campo Limpo Paulista (UNIFACCAMP), Campo Limpo Paulista, SP, Brasil*

[3] *Pesquisador, Centro de Tecnologia da Informação Renato Archer, Campinas, SP, Brasil*

[4] *Pesquisador, Centro de Tecnologia da Informação Renato Archer, Campinas, SP, Brasil*

[5] *Pesquisador, Centro de Tecnologia da Informação Renato Archer, Campinas, SP, Brasil e Professor, Centro Universitário Campo Limpo Paulista (UNIFACCAMP), Campo Limpo Paulista, SP, Brasil*

Autor Correspondente: **Rodrigo Bonacin**
e-mail: **rodrigo.bonacin@cti.gov.br**

Artigo recebido: 06/09/2022
Aprovado: 30/09/2023

1. INTRODUCTION

The evolution of Information Technology has made it possible to store large amounts of data in Electronic Health Records (EHRs). Although there is a recent evolution in EHR models and Knowledge Organization Systems (KOS), data quality, information recovery and semantics interoperability, issues still persist. Therefore, research for alternatives to provide EHR standard descriptions and annotations, considering both syntactic and semantic aspects, is required. International standards for EHR were developed aiming to provide interoperability and reuse, as well as to improve data quality of EHRs. However, such specifications are not sufficient from a semantic point of view, thus requiring the use of health domain-specific KOS.

In the biomedical field of research, there are several KOS that aim to represent the semantics of the domain. The increase in the number of biomedical KOS generates semantic integration and interoperability problems(1).

There are several reviews about usage (e.g.,(2)), specifications, and methods for EHRs development. Häyrinen et al.(3), for example, reviews the literature on EHR systems in order to analyze their use and impact. Yet Hwang et al.(4) reviewed the literature about the use of technologies to improve the interoperability of ubiquitous healthcare systems. Other reviews, such as Wollersheim et al.(5) and Frade et al.(6), explore different aspects of the OpenEHR implementation.

The use of different KOS has also been explored in the health informatics literature. Martins and Bulcão-Neto(7) present a review on methods for mapping ontologies to OpenEHR archetypes. There are reviews about the application of EHR in several health domains(8), such as medical education(9). However, there is a lack of reviews focusing on oncology, particularly those aimed to analyze the use of semantic technologies in EHR, as well as the use of KOS for pediatric oncology.

We aim to review the use of ontologies and other KOS as a means of enriching EHR for the oncology domain, with the focus on pediatric oncology. To provide a broad review on this topic, we carried out three reviews focused on: 1. semantic technologies in EHR for oncology (in general); 2. the use of KOS for pediatric oncology, and 3. a tertiary study aiming at Genetic Ontology, NCI Thesaurus, and pediatric oncology.

2. METHODOLOGY

The adopted methodology is based on Kitchenham(10) and Prisma(11). From results of an initial exploratory analysis of the literature and the research objectives, the following guide questions were elaborated: Q1. Which are the models and applied techniques aimed to deal with semantics in cancer-targeted or oncology EHRs? Q2. Which is the KOS applied to pediatric oncology studies? Q3. Which is the relation between genetic ontology KOS, or NCI thesaurus, and childhood cancer? The search strings presented in Table 1 were established from the proposed questions and respective relevant search terms, and synonyms of these terms.

Table 1 - Inclusion and Exclusion Criteria for Each Review

Search String 1	Search String 2	Search String 3
(“semantic” OR “ontology” OR “thesaurus”) AND (“electronic health record” OR “electronic medical record”) AND (“cancer” OR “oncology” OR “carcinoma” OR “tumor” OR “neoplasm” OR “cancerology”)	(“semantic” OR “ontology” OR “ontologies”) AND (“childhood cancer” OR “pediatric oncology”)	(“go ontology” OR “gene ontology” OR “National Cancer Institute”) AND (“cancer” OR “oncology” OR “carcinoma” OR “neoplasm”) AND (“childhood” OR “pediatric”)

The searches comprise articles published from 2016 to July/2023 indexed in: PubMed, Scopus, IEEE Xplore, and ACM Digital Library. Duplicate publications were excluded, titles and keywords were read in pairs, and the inclusion and exclusion criteria (Table 2) were applied. Then, the full text was peer-reviewed, and the criteria were applied again, selecting the most relevant results for the qualitative synthesis.

Table 2 - Inclusion and Exclusion Criteria for Each Review

Review	Inclusion Criterion	Exclusion Criterion
Rev. 1	• Complete work • Address EHR and Semantics in oncology • Primary and secondary studies • Published full paper	• Addressing only one of the topics • Less than five health records • Tertiary studies and protocol descriptions • Languages other than English or Portuguese
Rev. 2	• Complete work • Address ontologies and pediatric oncology • Primary and secondary studies • Published full paper	• Addressing only one of the topics • Less than five health records • Tertiary studies and Protocol descriptions • Languages other than English or Portuguese
Rev. 3	• Complete work • Addressing GO or NCI in pediatric oncology • Tertiary studies	• Not addressing pediatric oncology • Experience of less than five health records • Primary and secondary studies, and Protocols • Languages other than English or Portuguese

3. RESULTS

Figure 1 presents the quantitative results for each stage of the Prisma diagram. As presented below, there are 32 papers described in the first review, 17 in the second review, and 11 in the third review.

Table 3 presents the summary of the results of Review 1, including the use of the EHR standards, KOS adopted in the EHR and the key contributions. Table 4 presents the summary of Review 2, including KOS adopted in the EHR and the key contribution. Eleven studies were analyzed in Review 3, of each the studies(12,13,14,15,16) use GO, work(17) used both GO and NCIt, works(18, 19) refer to the NCIt, and other three works are undefined regarding the use of KOS.

Figure 1: Prisma diagram with quantitative results of reviews 1, 2 and 3

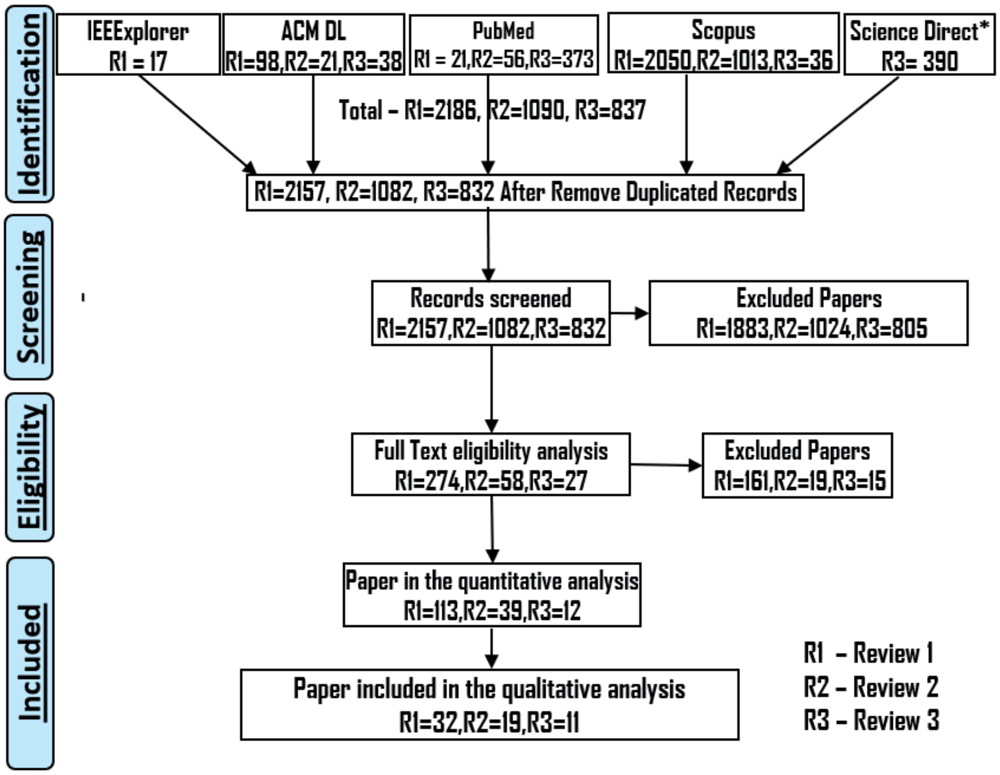

Table 3 – Summary table with the results from Review 1

Authors, year and reference	EHR standards	KOS in oncology EHRs	Type of Cancer	Key Contribution
Alkarkoukly & Rajput, 2021(20)	OpenEHR	ICD, ICD-O, TNM	Pancreas	Propose a virtual patient template
Banerjee et al., 2018(21)	Undefined	Semantic Mapping	General	Estimate the short-term life expectancy
Bibault et al., 2018(22)	Undefined	Bioportal, GO, OWL	Lymph Nodes	A new ontology specific to radiation oncology
Chan et al., 2016(23)	Undefined	SNOMED CT	Liver	Reveals association patterns in liver cancer
Chan et al., 2017(24)	Undefined	UMLS	Liver	Signifies liver cancer and decision support
Chang et al., 2019(25)	Undefined	Undefined	Breast	Hormone receptor status via data data mining
Chen et al., 2022(26)	Undefined	UMLS	Lung	Risk evaluation and preventive screening
Chiudinelli et al., 2020(27)	Undefined	ICD9	Breast	Provides evidence to support clinicians in post-surgical care
Dugas et al., 2016(28)	Undefined	SNOMED CT, UMLS	General	Development of uniform semantic annotations
Figueiredo et al., 2021(29)	OpenEHR	UMLS, SNOMED, NCI	General	Improved reuse and interop. of EHR models
Frid et al., 2022(30)	HL7	Various Ontologies	General	Integration patient app with clinical repository
Hanauer et al., 2020(31)	Undefined	UMLS	General	Improved search Engine for cancer research
He et al., 2019(32)	Undefined	UMLS, NCIt, Ont.	General	Improved maintenance of biomedical KOS
Hochheiser et al., 2016(33)	HL7	FHIR OWL, NCIt	General	A model for computable cancer phenotypes
Hong et al., 2018(34)	Undefined	ICD10, SNOMED, LOINC, NDFRT, READ	Liver/Cholangiocarcinoma	Design, development, and evaluation of generic survival analysis routines
Iatraki et al., 2018(35)	PHIR	Various Ontologies	General	Improved search of documents by patients

Koopman et al., 2018(36)	Undefined	SNOMED CT	Comm. & rare	Automatic cancer characterization
Kralj et al., 2019(37)	Undefined	Go ontology	LLA, Breast	Time efficiency improvements in rule learning
Legaz-García et al., 2016(38)	OpenEHR	OWL, Archetypes	Colorectal	Interoperability and exploitation of archetypes, EHR and ontologies
Li et al., 2020(39)	SEER Medicare	CCAE	General	A model for embedding categorical attributes used in clinical endpoint prediction
Maggi et al., 2019(40)	HL7	HL7-CDA	General	Management of patient data from wearables
Malty et al., 2018(41)	Undefined	RxNorm, OWL	Breast	Ontology of hematological drugs and regimens
Meng et al., 2018(42)	Undefined	SNOMED CT	Kidney	An early warning model based on EHRs
Messaoudi et al., 2019(43)	Undefined	Various Ontologies	Liver	Improved image analysis of liver lesions
Najafabadipour et al., 2020(44)	Undefined	Undefined	Lung	Extract natural history from EHR
Rani et al., 2019(45)	Undefined	SNOMED CT	Breast	Use of text annotations to support pathologists
Rubinstein et al., 2020(46)	Undefined	Vocabulary Scheme	Myeloma	Normalized chemotherapy nomenclature
Sharma et al., 2017(47)	OpenEHR	RDF, Archetypes	General	Facilitate clinical modeling in cancer genome
Sweidan et al., 2020(48)	Undefined	Various Ontologies	Liver	A fuzzy Ontology for liver fibrosis diagnosis
Wu et al., 2019(49)	VUMC HER	Undefined	General	Drug repurposing based on EHR data
Zexian et al., 2018(50)	Undefined	UMLS	Breast	Identify distant recurrences in breast cancer
Zong et al., 2021(51)	HL7	RDF Graphs	General	Prediction of primary cancers

4. DISCUSSION

Regarding Question Q1, results from Review 1 indicate that EHR standards are not used in most studies. Among the used standards, OpenEHR and HL7 excel from the others. This reveals that, although literature presents efforts to adopt these standards, there are still several projects that do not adopt international standards.

Table 4 - Summary table with the results from Review 2

Authors, year and reference	KOS in EHRs	Type of Cancer	Key Contribution
Chen et al, 2022b(52)	GO, KEGG	Neuroblastoma	Screened differential miRNAs related to NB
Dawidowska et al., 2019(53)	GO, KEGG	Acute Lymp. Leukemia	Identify mi-mRNA interaction implicated in T-Cell
Fonseka et al., 2019(54)	GO	Neuroblastoma	Role of exosomes in N-Myc-driven aggressive NB
Guo et al., 2021(55)	GO, KEGG	Papillary Thyroid	Biological mechanisms involved in CAYA-PTC
Huang et al., 2020(56)	GO	Medulloblastoma	Identify hub genes in pediatric medulloblastoma
Liu et al., 2017(57)	GO, KEGG e WGCNA	Osteosarcoma	Identify gene modules in human osteosarcoma
Luo et al., 2019(58)	GO, KEGG	Wilms Tumor	HnRNPL act as p53 mRNA-binding protein in WT
Mousavian et al., 2017(59)	GO	Mixed Lineage Leukemia	Reveals key genes related to glucocorticoid resistance in infant acute lymphoblastic leukemia
Núñez-Enríquez et al., 2016(60)	GO, KEGG	Acute Lymp. Leukemia	Identify relevant biomarkers for high-risk relapses
Ohmura et al., 2021(61)	GO	Ewing's Sarcoma	RRM2 as a biomarker and therapeutic target
Olsson et al., 2016(62)	GO	Neuroblastoma	Identify methylated genes in neuroblastoma
Shen et al., 2018(63)	GO	Osteosarcoma	Identify protein's role in the OS development
Sun et al., 2015(64)	GO	Osteosarcoma	Suggest a set of genes that may be involved in OS
Tomar et al., 2019(65)	GO, KEGG	Acute Lymp. Leukemia	Identify the most Variable Genes and Transcription
Xu et al., 2023(66)	GO	Rhabdomyosarcoma	Identify urinary biomarkers of rhabdomyosarcoma
Yan et al., 2018(67)	GO, KEGG	Ewing's Sarcoma	Identify genes and pathways in Ewing's sarcoma
Zhang et al., 2020(68)	GO, KEGG e WGCNA	Wilms Tumor	Understanding of the WT molecular mechanism
Zhong et al., 2018(69)	GO	Neuroblastoma	Identify gene-signature to predict NB prognosis
Zhou et al., 2019(12)	GO	Wilms Tumor	Identify potential biomarkers for Wilms tumor

Among the selected studies, we highlight the references(40,33,51,30) once they use the HL7. Reference(40) focuses on the semantic architecture for interaction with wearables and sensors. Reference(33) uses FHIR, which is the interoperable resource model of HL7, with an OWL representation for cancer phenotype modeled using NCI terms. References(47,38,29,20) use the OpenEHR standard. In(47) OpenEHR Reference Model is used, which is a set of standard classes and mini-archetype models. It aims to produce a repository of clinical cancer genome metadata, with application in a case study of the "pharmaceutical-clinical" domain in the TCGA. Article(38) describes an OWL file for clinical data, implemented as an archetype in the colorectal cancer domain. Works(16,49,35) address semantic EHR using lesser-known standards. The other works address EHRs but do not report the use of a well-defined standard.

As shown in Table 3, SNOMED CT is the most cited alone and together with other KOS, including UMLS, NCIt, and ICD. Unlike the use of international EHR standards, few articles do not mention the use of at least a KOS. This may show higher attention to the term and concept descriptions than to the use of standard formats for EHRs. Thus, the use of EHR standards with recent technologies in the field of semantic interoperability is a key point that is still in a future research agenda. In(45), SNOMED CT is used as an annotation standard for a Clinical Decision Support System (CDSS) in the realm of breast cancer. CDSS supports cancer pathological classification and staging with natural language annotation. In(36), a classifier is proposed to identify cancer-related causes of death; this classifier is used to extract SNOMED CT concepts. In(42), the authors use EHR history and SNOMED CT semantic relationships to expand terms related to kidney cancer, to achieve higher diagnostic accuracy.

In(23), SNOMED CT is used to analyze features of the liver. In(34), six KOS (ICD10CM, ICD10PCS, SNOMED, LOINC, NDFRT, and READ) are used to support survival analysis in intrahepatic cholangiocarcinoma patients. In(28), UMLS and SNOMED CT are used in the development of semantic annotations for medical data, aiming to provide uniform codes. In(33) cancer phenotypes are modeled with terms from the NCI Thesaurus, while(32) describes a cross-ontology topological standard, using UMLS and NCIt. In(27), the authors use ICD9 in a care flow algorithm based on the tumor's clinical and histopathological characteristics. As shown in Table 3, several works do not focus on a specific cancer. Breast cancer is the most studied one, followed by liver cancer. Various other cancer types are listed in one work each.

Regarding Question Q2, results from Review 2 (Table 4) show that Gene Ontology (GO) is the most used one, including its use alone and with KEGG (Kyoto Encyclopedia of Genes and Genomes) and WGCNA (Weighted Gene Coexpression Network Analysis), which are both linked to genetics. This result differs from the KOS pointed out in our first review.

As shown in Table 4, when considering childhood cancer types, the following stand out: Acute Lymphoblastic Leukemia (ALL) and Wilm's Tumor, followed by Neuroblastoma and Osteosarcoma. Medulloblastoma, Ewing's Sarcoma, and Retinoblastoma are present in one study. These have characteristics related to genetics and heredity or are related to the blood and immune system. This may justify the large number of works related to GO combined with other KOS focused on the genetics domain, as well as their integration with other KOS such as NCIt.

In(57), the WGCNA is used to build a genetic co-expression network for the prediction of involved gene clusters in osteosarcoma. In(68), the GO, KEGG, and WGCNA are used to analyze cell proliferation and ways to assess Wilm's Tumors progression and to develop targeted therapies and biomarkers. In(60), the KEGG and GO are used to identify new biomarkers and to analyze ALL's gene expression.

In(53), eight algorithms that exploit GO and KEGG were used in order to analyze miRNA-mRNA interactions for ALL. In(67), GEO (Gene Expression Omnibus) and DAVID (Database for Annotation, Visualization, and Integrated Discovery) databases are used in conjunction with GO and KEGG to identify biomarkers and the process of Ewing's Sarcoma. In(58), GO and KEGG are used to discover differentially expressed genes between Wilm's Tumor and non-adjacent tumors. In(12), the authors make use of GO to discover an endogenous competitor RNA and a biomarker of Wilm's Tumor.

In(65), a mRNA expression analysis is performed to annotate genes and aggression of ALL, GO and KEGG are used. In(69), using GEO and DAVID bases with GO, a gene signature is identified in order to predict the prognosis of patients with neuroblastoma. In(53), GO is applied in a methylation profile analysis, which can be used in patient stratification and inform about epigenetically dysregulated genes. GO is used in(54), where the authors point out that exosomes play a key role in aggressive neuroblastoma driven by amplification of the oncogene and in chemoresistance transfer between cells. In(59), GO is used in a weighted genetic co-expression network to identify gene modules associated with GC glucocorticoids in pediatric patients with ALL. In(63), is examined the molecular mechanisms of osteosarcoma. The GO is used to analyze the potential functions of Differentially Expressed Genes (DEGs). In(56), GO is used in the analysis of three microarray datasets. Reference(64) investigates the osteosarcoma; DEGs were analyzed by using GO.

Regarding Question Q3, the results from Revision 3 show that 45% of studies use GO, 18% NCIt, and 9% NCIt + GO together; the remaining (28%) is undefined. From a technological point of view, three of the analyzed works address languages and tools linked to machine learning for the prediction and identification of relevant information. However, we highlight that these studies neither address the link of these KOS with EHRs for pediatric oncology nor the integration with other KOS in EHRs.

In(17) big data is used to outline a treatment perspective in pediatric oncology using NCIt and GO. In(18), population-based cancer registries from Canada are used to standardize data collection, diagnosis, and out-of-stage prognosis. In(70), the authors address the TNM (Tumor Node Metastasis) classification, which is applied to adults, and aim to establish principles and standardization for TNM classification in pediatric oncology. In(19), the authors employed NCIt to analyze the survival rates in children with ALL and the treatment of acute toxic effects.

In(12), the authors analyze endogenous, and biomarkers related to Wilm's Tumor using GO. In(13), GO is used to annotate lncRNAs in a method aimed to predict large-scale human lncRNAs functions. In(14), a prediction model was developed for the prognosis of ALL cancer using clinical data and gene expression annotated with GO. In(71), the authors evaluate the performance and robustness of Optimal Bayesian Filtering for biomarker discovery in the context of pediatric oncology. In reference(18) a modified Delphi approach is used to guide the Pediatric Cancer Data gathering. Work(15) uses GO in the visualization of genomic characteristics and reference(16) uses GO to reveal recurrent genetic and transcriptomic signatures.

This work does not analyze in depth the benefits of adopting each standard and KOS from a system and clinical point of view. Therefore, a systematic long-term study is necessary. However, this work already provides valuable insights into the adoption of semantic technologies in EHR and the use of KOS in pediatric oncology.

5. CONCLUSION

Although there are advances in the use of international EHR specification standards as well as advances of KOS of the health domain, semantic interoperability, model reuse, and data quality are still challenges to be faced on health informatics research. We presented an integrative review, composed of three literature reviews. We identified and analyzed the research and use of standards and KOS for semantic specification of EHRs in the realm of pediatric oncology.

Our literature review points out that the most frequently used standards in the analyzed works are HL7 and OpenEHR. However, many studies do not adopt a specific standard. The review also points out differences between the main KOS used in oncology in general and the most used KOS in the domain of pediatric oncology. From the results, the need for further research and improvement in the specification of EHRs that integrate KOS with international standards was identified. In the next steps, we expect to propose processes and models aimed at integrating standards for EHRs specifications with the most used KOS for pediatric oncology domain.

ACKNOWLEDGEMENTS

This study was financed in part by The Brazilian National Council for Scientific and Technological Development (CNPq) grant 405352/2021-2 and São Paulo Research Foundation (FAPESP) grant 2021/11905-0.

REFERENCES

1. Farinelli F, Almeida MB. Semantic interoperability in healthcare information systems: A top-level ontology based solution. In: *4th Enterprise Engineering Working Conference (EEWC 2014)*, Doctoral consortium; Funchal, Madeira, Portugal; 2014.
2. Kruse CS, Stein A, Thomas H, Kaur H. The use of electronic health records to support population health: A systematic review of the literature. *J Med Syst.* 2018;42(11):214. DOI: 10.1007/s10916-018-1075-6.
3. Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: A review of the research literature. *Int J Med Inform.* 2008;77(5):291-304. DOI: 10.1016/j.ijmedinf.2007.09.001.
4. Hwang KH, Chung KI, Chung MA, Choi D. Review of semantically interoperable electronic health records for ubiquitous healthcare. *Healthc Inform Res.* 2010;16(1):1-5. DOI: 10.4258/hir.2010.16.1.1.
5. Wollersheim D, Sari A, Rahayu W. Archetype-based electronic health records: A literature review and evaluation of their applicability to health data interoperability and access. *Health Inf Manag J.* 2009;38(2):7-17. DOI: 10.1177/183335830903800202.
6. Frade S, Freire SM, Sundvall E, Patriarca-Almeida JH, Cruz-Correia R. Survey of openEHR storage implementations. In: *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems.* 2013; pp. 303-307. IEEE. DOI: 10.1109/CBMS.2013.6627806.
7. Martins MC, Bulcão-Neto RF. Métodos de mapeamento entre arquétipos e ontologias: Uma revisão sistemática. *J Health Informatics.* 2019;11(1). URL: http://www.jhi-sbis.saude.ws/ojs-jhi/index.php/jhi-sbis/article/view/634.
8. Caetano CS, Orfão NH. Prontuário eletrônico do paciente em clínicas odontológicas: Uma revisão integrativa. *J Health Informatics*. 2021;13(3). URL: http://www.jhi-sbis.saude.ws/ojs-jhi/index.php/jhi-sbis/article/view/860.
9. Rangel AMP, Struchiner M. Prontuário eletrônico do paciente (pep) na educação médica: Benefícios e preocupações. *J Health Informatics*. 2021;13(2). URL: http://www.jhi-sbis.saude.ws/ojs-jhi/index.php/jhi-sbis/article/view/798.
10. Kitchenham B. Procedures for performing systematic reviews. *Joint Technical Report TR/SE-0401 and 0400011T.1*, Keele University and National ICT Australia Ltd., 2004.
11. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care

interventions: Explanation and elaboration. *J Clin Epidemiol.* 2009;62(10):e1-e34. DOI: 10.1016/j.jclinepi.2009.06.006.

12. Zhou Y, Liang H, Deng J. Bioinformatics analysis of competing endogenous RNA and screening of biomarkers in childhood Wilms tumor. In: *Proceedings of the 2019 4th International Conference on Intelligent Information Processing*, ICIIP 2019, p. 486-490. Association for Computing Machinery, New York, NY, USA. DOI: 10.1145/3378065.3378156.
13. Zhang Z, Zhang J, Fan C, Tang Y, Deng L. KatzLGO: Large-scale prediction of lncRNA functions by using the Katz measure based on multiple networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. 2019;16(2):407-416. DOI: 10.1109/TCBB.2017.2704587.
14. Ghaisani FD, Wasito I, Faturrahman M, Mufidah R. Deep belief networks and bayesian networks for prognosis of acute lymphoblastic leukemia. In: *Proceedings of the International Conference on Algorithms, Computing and Systems*, ICACS '17, 2017; p. 102–106. Association for Computing Machinery, New York, NY, USA. DOI: 10.1145/3127942.3127947.
15. Arora S, Szulzewsky F, Jensen M, Nuechterlein N, Pattwell SS, Holland EC. Visualizing genomic characteristics across an RNA-Seq based reference landscape of normal and neoplastic brain. *Sci Rep.* 2023;13:4228. https://doi.org/10.1038/s41598-023-31180-z
16. Lazzarini E, Silvestris DA, Benvenuto G, Osti D, Fattore L, Paterra R, et al. Genome-wide profiling of patient-derived glioblastoma stem-like cells reveals recurrent genetic and transcriptomic signatures associated with brain tumors. *J Neurooncol.* 2023 May;163(1):47-59. DOI: 10.1007/s11060-023-04287-6.
17. Major A, Cox SM, Volchenboum SL. Using big data in pediatric oncology: Current applications and future directions. *Seminars in Oncology*. 2020;47(1):56-64. DOI: https://doi.org/10.1053/j.seminoncol.2020.02.006.
18. Gupta S, Aitken JF, Bartels U, Brierley J, Dolendo M, Friedrich P, et al. Development of pediatric non--stage prognosticator guidelines for population-based cancer registries and updates to the 2014 Toronto pediatric cancer stage guidelines. *Lancet Oncol.* 2020;21(9):e444-e451.
19. Schmiegelow K, Attarbaschi A, Barzilai S, Escherich G, Frandsen TL, Halsey C, et al. Ponte di Legno Toxicity Working Group: Consensus definitions of 14 severe acute toxic effects for childhood lymphoblastic leukaemia treatment: A Delphi consensus. *Lancet Oncol.* 2016;17(6):e231-e239.
20. Alkarkoukly S, Rajput AM. An openEHR Virtual Patient Template for Pancreatic Cancer. *Stud Health Technol Inform.* 2021;285:292–295. https://doi.org/10.3233/SHTI210618
21. Banerjee I, Gensheimer MF, Wood DJ, Henry S, Aggarwal S, Chang DT, Rubin DL. Probabilistic Prognostic Estimates of Survival in Metastatic Cancer Patients (PPES-Met) Utilizing Free-Text Clinical Narratives. *Scientific Reports.* 2018;8(1). https://doi.org/10.1038/s41598-018-27946-5
22. Bibault J-E, Zapletal E, Rance B, Giraud P, Burgun A. Labeling for big data in radiation oncology: The radiation oncology structures ontology. *PLoS ONE.* 2018;13(1). https://doi.org/10.1371/journal.pone.0191263
23. Chan LW, Wong SC, Chiu KW. Ontological features of electronic health records reveal distinct association patterns in liver cancer. In: *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).* 2016; pp. 1051-1053. DOI: 10.1109/BIBM.2016.7822667.
24. Chan LWC, Cesar Wong SC, Chiau CC, Chan T-M, Tao L, Feng J, Chiu KWH. Association Patterns of Ontological Features Signify Electronic Health Records in Liver Cancer. *Journal of Healthcare Engineering.* 2017. https://doi.org/10.1155/2017/6493016
25. Chang KP, Chu YW, Wang J. Analysis of hormone receptor status in primary and recurrent breast cancer via data mining pathology reports. *Open Med.* 2019;14(1):91-98. DOI: doi:10.1515/med-2019-0013.
26. Chen A, Huang R, Wu E, Han R, Wen J, Li Q, Zhang Z, Shen B. The Generation of a Lung Cancer Health Factor Distribution Using Patient Graphs Constructed From Electronic Medical Records: Retrospective Study. *J Med Internet Res.* 2022;24(11):e40361. https://doi.org/10.2196/40361
27. Chiudinelli L, Dagliati A, Tibollo V, Albasini S, Geifman N, Peek N, et al. Mining post-surgical care processes in breast cancer patients. *Artif Intell Med.* 2020;105:101855. DOI: https://doi.org/10.1016/j.artmed.2020.101855.
28. Dugas M, Meidt A, Neuhaus P, Storck M, Varghese J. OdMediT: Uniform semantic annotation for data integration in medicine based on a public metadata repository. *BMC Med Res Methodol.* 2016;16(1):65. DOI: 10.1186/s12874-016-0164-9.
29. de Figueiredo EB, Dametto M, Rosa FF, Bonacin R. A Multidimensional Framework for Semantic Electronic Health Records in the Oncology Domain. In: *2021 IEEE 30th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, Bayonne, France; 2021. pp. 165-170. https://doi.org/10.1109/WETICE53228.2021.00041
30. Frid S, Expósito MA, Grau-Corral I, Amat-Fernandez C, Mateu M, Duran X, Lozano-Rubí R. Successful Integration of EN/ISO 13606-Standardized Extracts From a Patient Mobile App Into an Electronic Health Record: Description of a Methodology. *JMIR Med Inform.* 2022;10(10):e40344. https://doi.org/10.2196/40344

31. Hanauer DA, Barnholtz-Sloan JS, Beno MF, Del Fiol G, Durbin EB, Gologorskaya O, Harris D, Harnett B, Kawamoto K, May B, Meeks E, Pfaff E, Weiss J, Zheng K. Electronic Medical Record Search Engine (EMERSE): An Information Retrieval Tool for Supporting Cancer Research. *JCO Clinical Cancer Informatics.* 2020;4:454–463. https://doi.org/10.1200/CCI.19.00134
32. He Z, Keloth VK, Chen Y, Geller J. Extended analysis of topological-pattern based ontology enrichment. In: *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).* 2018; pp. 1641-1648. DOI: 10.1109/BIBM.2018.
33. Hochheiser H, Castine M, Harris D, Savova G, Jacobson RS. An information model for computable cancer phenotypes. *BMC Med Inform Decis Mak.* 2016;16(1):121.DOI:10.1186/s12911-016-0358-4.
34. Hong N, Zhang N, Wu H, Lu S, Yu Y, Hou L, et al. Preliminary exploration of survival analysis using the OHDSI common data model: a case study of intrahepatic cholangiocarcinoma. BMC *Med Inform Decis Mak.* 2018;18(5):116. DOI: 10.1186/s12911-018-0686-7.
35. Iatraki G, Kondylakis H, Koumakis L, Chatzimina M, Kazantzaki E, Marias K, et al. Personal health information recommender: implementing a tool for the empowerment of cancer patients. *Ecancermedicalscience.* 2018;12:851.
36. Koopman B, Zuccon G, Nguyen A, Bergheim A, Grayson N. Extracting cancer mortality statistics from death certificates: A hybrid machine learning and rule-based approach for common and rare cancers. *Artif Intell Med.* 2018;89:1-9. DOI: https://doi.org/10.1016/j.artmed.2018.04.011.
37. Kralj J, Robnik-Šikonja M, Lavrač N. NetSDM: Semantic data mining with network analysis. *J Mach Learn Res.* 2019;20(1):1086-1135.
38. Legaz-García MC, 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. *Knowledge-Based Systems.* 2016;105:175-189. DOI: https://doi.org/10.1016/j.knosys.2016.05.016.
39. Li Y, Zhu Z, Wu H, Ding S, Zhao Y. CCAE: Cross-field categorical attributes embedding for cancer clinical endpoint prediction. *Artif Intell Med.* 2020;107:101915. DOI: https://doi.org/10.1016/j.artmed.2020.101915.
40. Maggi N, Magnoni LD, Ruggiero C, Gazzarata R, Giacomini M. Information technology system including patient-generated health data for cancer clinical care and research. *Stud Health Technol Inform.* 2019;261:289-293.
41. Malty AM, Jain SK, Yang PC, Harvey K, Warner JL. Computerized approach to creating a systematic ontology of hematology/oncology regimens. *JCO Clin Cancer Informatics.* 2018;(2):1-11. DOI: 10.1200/CCI.17.00142.
42. Meng J, Zhang R, Chen D. Utilizing narrative text from electronic health records for early warning model of chronic disease. In: *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC*). 2018; pp. 4210-4215. DOI: 10.1109/SMC.2018.00713.
43. Messaoudi R, Jaziri F, Mtibaa A, Grand-Brochier M, Ali HM, Amouri A, et al. Ontology-based approach for liver cancer diagnosis and treatment. *J Digit Imaging.* 2019;32(1):116-130. DOI: 10.1007/s10278-018-0115-6.
44. Najafabadipour M, Zanin M, Rodríguez-González A, Torrente M, Nuñez García B, Cruz Bermudez JL, Provencio M, Menasalvas E. Reconstructing the patient's natural history from electronic health records. *Artificial Intelligence in Medicine.* 2020;105. https://doi.org/10.1016/j.artmed.2020.101860
45. Rani GJJ, Gladis D, Mammen JJ. Snomed ct annotation for improved pathological decisions in breast cancer domain. *Int J Recent Technol Eng.* 2019;8(3):8400-8406.
46. Rubinstein SM, Yang PC, Cowan AJ, Warner JL. Standardizing Chemotherapy Regimen Nomenclature: A Proposal and Evaluation of the HemOnc and National Cancer Institute Thesaurus Regimen Content. *JCO Clinical Cancer Informatics.* 2020;3(4):60–70. https://doi.org/10.1200/CCI.19.00122
47. Sharma DK, Solbrig HR, Tao C, Weng C, Chute CG, Jiang G. Building a semantic web-based metadata repository for facilitating detailed clinical modeling in cancer genome studies. *J Biomed Semantics.* 2017;8(1):19. DOI: 10.1186/s13326-017-0130-4.
48. Sweidan S, El-Bakry H, Sabbeh SF. Construction of liver fibrosis diagnosis ontology from fuzzy extended ER modeling: Construction of Fibron-To from an EER model. *Int J Decision Support Syst Technol (IJDSST).* 2020;12(1):46-69. DOI: 10.4018/IJDSST.2020010103.
49. Wu Y, Warner JL, Wang L, Jiang M, Xu J, Chen Q, et al. Discovery of noncancer drug effects on survival in electronic health records of patients with cancer: A new paradigm for drug repurposing. *JCO Clin Cancer Inform.* 2019;(3):1-9. DOI: 10.1200/CCI.19.00001. PMID: 31141421.
50. Zexian Z, Ankita R, Xiaoyu L, Sasa E, Susan C, Seema K, et al. Using clinical narratives and structured data to identify distant recurrences in breast cancer. In: *2018 IEEE International Conference on Healthcare Informatics (ICHI).* 2018; pp. 44-52. DOI: 10.1109/ICHI.2018.00013.
51. Zong N, Ngo V, Stone DJ, Wen A, Zhao Y, Yu Y, Liu S, Huang M, Wang C, Jiang G. Leveraging Genetic Reports and Electronic Health Records for

the Prediction of Primary Cancers: Algorithm Development and Validation Study. *JMIR Med Inform.* 2021;9(5):e23586. https://doi.org/10.2196/23586

52. Chen ZF, Wu LZ, Chen ZT, Su LJ, Fu CJ. The potential mechanisms of neuroblastoma in children based on bioinformatics big data. *Transl Pediatr.* 2022;11(12):1908–1919. https://doi.org/10.21037/tp-22-504
53. Dawidowska M, Jaksik R, Drobna M, Szarzyńska-Zawadzka B, Kosmalska M, Łukasz Sędek, et al. Comprehensive investigation of mirnome identifies novel candidate miRNA-mRNA interactions implicated in T-cell acute lymphoblastic leukemia. *Neoplasia.* 2019;21(3):294-310. DOI: https://doi.org/10.1016/j.neo.2019.01.004.
54. Fonseka P, Liem M, Ozcitti C, Adda CG, Ang CS, Mathivanan S. Exosomes from N-Myc amplified neuroblastoma cells induce migration and confer chemoresistance to non-N-Myc amplified cells: Implications of intra-tumour heterogeneity. *J Extracell Vesicles.* 2019;8(1):1597614. DOI: https://doi.org/10.1080/20013078.2019.1597614.
55. Guo K, Qian K, Shi Y, Sun T, Chen L, Mei D, Dong K, Gu S, Liu J, Lv Z, Wang Z. Clinical and Molecular Characterizations of Papillary Thyroid Cancer in Children and Young Adults: A Multicenter Retrospective Study. *Thyroid.* 2021;31(11):1693–1706. https://doi.org/10.1089/thy.2021.0003
56. Huang P, Guo YD, Zhang HW. Identification of hub genes in pediatric medulloblastoma by multiple-microarray analysis. *Journal of Molecular Neuroscience.* 2020; 70(4), 522–531. DOI: https://doi.org/10.1007/s12031-019-01451-4.
57. Liu X, Hu A.X., Zhao J.L., Chen F.L. Identification of key gene modules in human osteosarcoma by co-expression analysis weighted gene co-expression network analysis (WGCNA). *Journal of Cellular Biochemistry*. 2017; 118(11), 3953–3959. DOI: https://doi.org/10.1002/jcb.26050.
58. Luo X, Deng C, Liu F, Liu X, Lin T, He D, et al. hnRNPL promotes Wilms tumor progression by regulating the p53 and Bcl2 pathways. *Onco Targets Ther.* 2019;12:4269-4279.
59. Mousavian Z, Nowzari-Dalini A, Stam R, Rahmatallah Y, Masoudi-Nejad A. Network-based expression analysis reveals key genes related to glucocorticoid resistance in infant acute lymphoblastic leukemia. *Cell Oncol.* 2017;40(1):33-45. DOI: 10.1007/s13402-016-0303-7.
60. Núñez-Enríquez JC, Bárcenas-López DA, Hidalgo-Miranda A, Jiménez-Hernández E, Bekker-Méndez VC, Flores-Lujano J, et al. Gene expression profiling of acute lymphoblastic leukemia in children with very early relapse. *Arch Med Res.* 2020;47(8):644-655. DOI: https://doi.org/10.1016/j.arcmed.2016.12.005.
61. Ohmura S, Marchetto A, Orth MF, et al. Translational evidence for RRM2 as a prognostic biomarker and therapeutic target in Ewing sarcoma. *Mol Cancer.* 2021;20:97. https://doi.org/10.1186/s12943-021-01393-9
62. Olsson M, Beck S, Kogner P, Martinsson T, Carén H. Genome-wide methylation profiling identifies novel methylated genes in neuroblastoma tumors. *Epigenetics.* 2016;11(1):74–84. https://doi.org/10.1080/15592294.2016.1138195
63. Shen H, Wang W, Ni B, Zou Q, Lu H, Wang Z. Exploring the molecular mechanisms of osteosarcoma by the integrated analysis of mRNAs and miRNA microarrays. *Int J Mol Med.* 2018;42(1):21-30. DOI: 10.3892/ijmm.2018.3594.
64. Sun L, Li J, Yan B. Gene expression profiling analysis of osteosarcoma cell lines. *Mol Med Rep.* 2015;12(3):4266-4272. DOI: 10.3892/mmr.2015.3958.
65. Tomar AK, Agarwal R, Kundu B. Most variable genes and transcription factors in acute lymphoblastic leukemia patients. *Interdiscip Sci.* 2019;11(4):668-678. DOI: 10.1007/s12539-019-00325-y.
66. Xu N, Yu Y, Duan C, Wei J, Sun W, Jiang C, Jian B, Cao W, Jia L, Ma X. Quantitative proteomics identifies and validates urinary biomarkers of rhabdomyosarcoma in children. *Clin Proteomics.* 2023;20:10. https://doi.org/10.1186/s12014-023-09401-4
67. Yan C, Wang Y, Wang Q, Feng X, Wang L, Bu Z, et al. Identification of key genes and pathways in Ewing's sarcoma using bioinformatics analysis. *J BUON.* 2018;23(5):1472-1480.
68. Zhang F, Zeng L, Cai Q, Xu Z, Liu R, Zhong H, et al. Comprehensive analysis of a long noncoding RNA-associated competing endogenous RNA network in Wilms tumor. *Cancer Control.* 2020;27(2). DOI: 10.1177/1073274820936991.
69. Zhong X, Liu Y, Liu H, Zhang Y, Wang L, Zhang H. Identification of potential prognostic genes for neuroblastoma. *Front Genet.* 2018;9:589.
70. Gupta S, Aitken JF, Bartels U, Brierley JD, Dolendo M, Friedrich P, et al. Pediatric cancer stage in population-based cancer registries: The Toronto consensus principles and guidelines. *Lancet Oncol.* 2016;17(4):e163-e172.
71. Foroughi Pour A, Dalton LA. Optimal Bayesian filtering for biomarker discovery: Performance and robustness. *IEEE/ACM Transactions on Computational Biology and Bioinformatics.* 2020;17(1):250-263. DOI: 10.1109/TCBB.2018.2858814.